We are thus able to assert, with greater confidence than before, since it is not necessary to split up the 2 degrees of freedom, that the food effect is significant. Correcting the three mean growth rates for variable initial weight by subtracting $b(w_0 - \bar{w}_0)$, where b is the regression coefficient calculated from the error term of Table V, while w_0 is now the *mean* initial weight for treatments A, B or C, $\bar{w}_0$ being the general mean, we may summarise the results as follows:

TABLE VI

Summary of results—Corrected growth rate

Food treatment	A	B	C	Mean	Standard error
Lb. per week %	9·676 104·0	9·235 99·3	9·003 96·8	9·304 100·0	0·1592 1·71

The percentage drop from A to C is unaltered (for in fact the initial weights for these two groups were the same), but the standard error of the three figures is reduced from 2·11 to 1·71, which accounts for the greater significance of the fall.

It will be noted that the gilts are lighter in initial weight than the hogs, but have the higher growth rate, though neither effect is significant. In view, however, of the positive correlation between growth rate and initial weight (from Table V b is 0·0889, corresponding to an r of 0·649), it is of interest to examine whether a significant sex difference emerges after correction for initial weight. The test is as follows, utilizing Table V:

TABLE Vo

Analysis of residual variance—Sex

	D.F.	Sum of squares	Mean square
Sex+error Error	20 19	6·0749 4·8155	 0·2534
Difference	1	1·2594	1·2594 $z=0·8017$ *S*

The sex effect is now significant at the 5% level, and correcting for initial weight the mean growth rates for hogs and gilts separately we have the following result.

TABLE VI A

Summary of results—Corrected growth rate

Sex	Hogs	Gilts	Mean	Standard error
Lb. per week	9·092	9·517	9·304	0·1300
%	97·7	102·3	100·0	1 40

The difference between the growth rates for hogs and gilts is 0·425 lb. per week in favour of the gilts, which difference has a standard error of 0·1903, calculated as the square root of

$$0{\cdot}2534\left(\frac{2}{15}+\frac{2{\cdot}06^2}{442{\cdot}93}\right).$$

0·2534 is the error mean square residual, while we are examining the difference between two means of fifteen pigs each. 2·06 is the mean difference in initial weight between hogs and gilts, while 442·93 is the error sum of squares for initial weight from Table V.* On a percentage basis, the difference between hogs and gilts is 4·6, with a standard error of 2·05. The experiment was not specifically designed to examine sex differences in the growth of the pigs, nor is it known how far such differences in growth rate during what is after all only the early part of the normal pig's life (though it is the whole of the life of the pig destined for the bacon factory) are matters of common knowledge. Nevertheless there seems to be little doubt about the effect in the present case.

If pen differences are examined in the same way, it will be found that the residual mean square, after correcting for initial weight, is not significant ($z=0{\cdot}4225$, $n_1=4$, $n_2=19$). This confirms the view that the significant pen differences in growth rate are a consequence of the very different average initial weights at which the different litters entered the experiment, and there is no evidence that rate of growth is a litter characteristic of any particular significance.

ANALYSIS OF RATE OF CHANGE OF GROWTH RATE

A similar analysis to that of growth rate may be carried out on the parabolic term h of the curve fitted to the weight measures. The analysis of variance is shown in Table VII. It is clear, on examination of this table, that the only significant effect is that of sex. This is shown in Table VIII.

* See Wishart and Sanders (1935, p. 54).

TABLE VII

Analysis of variance of rate of change of growth rate

Variation due to	D.F.	Sum of squares	Mean square		
Pens	4	0·0034835	0·0008709		
Food	2	0·0012578	0·0006289		
Sex	1	0·0030603	0·0030603	$z = 0·8060$	*S*
Interaction	2	0·0008078	0·0004039		
Error	20	0·0122093	0·0006105		
Total	29	0·0208187			

Standard error per pig = $\sqrt{(0·0006105)}$ = 0·02686, or 15·63 % of the mean, 0·1719.

TABLE VIII

Summary of results—Rate of change of growth rate

Sex	Hogs	Gilts	Mean	Standard error
$\frac{1}{2}$ {lb./(week)^2}	0·1618	0·1820	0·1719	0·00638
%	94·1	105·9	100·0	3·71

Not only have the gilts shown a higher average growth rate than the hogs (when corrected for initial weight), but they now show a higher rate of change of growth rate, i.e. there is a greater degree of curvature in the growth figures. The difference in favour of the gilts is 11·8 %, with a standard error of 5·25.

Finally we may examine the rate of change figures in relation to initial weight. The table is as follows:

TABLE IX

Analysis of variance and covariance. Initial weight and rate of change of growth rate

Variation due to	D.F.	(w_0^2)	$(w_0 h)$	(h^2)	$b = (w_0 h)/(w_0^2)$	$(w_0 h)^2/(w_0^2)$
Pens	4	605·87	−0·18386	0·0034835		
Food	2	5·40	0·0117	0·0012578		
Sex	1	32·03	−0·3131	0·0030603		
Interaction	2	22·47	−0·1293	0·0008078		
Error	20	442·93	0·10186	0·0122093	0·00023	0·0000234
Total	29	1108·70	−0·5127	0·0208187		

$r_{w_0 h} = 0·0438$ *NS*

That the regression, however, of rate of change of growth rate on initial weight is not significant is shown by the following test:

TABLE IX A

Test of regression

Variation due to	D.F.	Sum of squares	Mean square
Regression	1	0·0000234	0·0000234 *NS*
Deviations	19	0·0121859	0·0006414

This being so, we are not likely to add to the information already obtained by examining the various effects when corrected for initial weight. No improvement, for example, is shown in the significance of the sex comparison. The downward trend shown in the figures of rate of change of growth rate with increasing protein percentage in the ration, while suggestive of what may be happening, is definitely not significant, even if the principal effect, with 1 degree of freedom, be isolated from the remainder.

DISCUSSION

By considering the actual weekly figures of the weights of the thirty pigs given over to this nutrition experiment, we have been able to demonstrate the significance of the fall in average growth rate with increasing protein percentage in the ration, and the sex difference in favour of the gilts in the rate of change of the growth rate, without making any allowance for initial weight. This contrasts with the previous study where only live-weight gain was considered. When the figures of mean growth rate are corrected for initial weight, the significance of the food effect is stronger, and a sex difference in favour of the gilts emerges as significant. Not only is the taking into consideration of the initial weights valuable from the point of view of reaching such conclusions, but it seems to be necessary to do so if we are to disentangle the sex comparison from the heavy-light comparison with which it is to some extent confounded by the design adopted for the experiment.

Were the decisions reached by separate examination of the growth rate and change of growth rate figures not so clear-cut, it might be necessary to take these figures (g and h of Table II) together in a simultaneous analysis of variance and covariance, and reach a single test of significance of the effect of food (or of sex) on both simultaneously, after the manner suggested by Bartlett (1934). The method outlined in this paper of calculating a number of quantities to express the growth of the pigs would seem, in fact, to be well adapted to this method of analysis, since we are seeking the effect of the food ration on growth, which is expressed by

both of the variables g and h (and possibly by the cubic term as well). Not only so, but the fact that it is desirable to take initial weights into account suggests that Bartlett's method should be applied to the partial variables derived from g and h when w_0 is held constant, and a test of significance derived in the same sort of way as in the usual covariance analysis. We have, in fact, a case of multiple dependent variables, with one independent variable, a special case of the kind envisaged by Day & Fisher (1937). This point is not pursued in the present paper, but is commended to the attention of investigators.

REFERENCES

Aitken, A. C. (1933). *Proc. Roy. Soc. Edinb.* **53**, 54–78.

Bartlett, M. S. (1934). *Proc. Camb. Phil. Soc.* **30**, 327–40.

Day, B. & Fisher, R. A. (1937). *Ann. Eugen., Camb.*, **7**, 333–48.

Fisher, R. A. (1936). *Statistical Methods for Research Workers*, Sixth Edn. § 27.

Wishart, J. & Sanders, H. G. (1935). *Principles and Practice of Field Experimentation*, p. 45 (Empire Cotton Growing Corporation).

Woodman, Evans, Callow & Wishart (1936). *J. Agric. Sci.* **26**, 546–619.

CERTAIN STATISTICAL PROBLEMS ARISING IN PLANT BREEDING

By Y. TANG

Department of Statistics, University College, London.

CONTENTS

I. Introduction

One of the important problems in agricultural science is the breeding and selection of new families or varieties which, for some economic reasons, are better than those already known. The desired properties of the plants are usually very complex and include a combination of various characters, yielding capacity, resistance to diseases, etc. However, to simplify the problem, we shall assume below that there is just one single character in plants, the importance of which is overwhelming and which it is desired to better by breeding new varieties.

The process of breeding new varieties depends on various circumstances, such as whether the plant under consideration is self-fertilizing or not. In the following I shall consider problems arising in the breeding of sugar beet, with a view to increasing their sugar content. It seems probable that similar problems are also met with in many other cases. It will be useful to call attention to two properties of sugar beet: (1) Sugar beet is a cross-fertilizing plant, which makes it practically impossible to obtain anything like a pure line. (2) The vegetation period of sugar beet covers 2 years. During the first year, a seedling produces a root rich in sugar but no seeds. The seeds are produced during the second year of life of the plant, when sugar stored in the root is used as foodstuff.

Before I describe the problem to be dealt with below, it will be useful to give some idea of the process of breeding. This is roughly explained in Fig. 1, which,

however, omits certain details and devices which are used by particular breeders and are not relevant from the point of view of the problems I am going to treat.

Fig. 1 shows the subdivision of the process of breeding into five steps. First certain individual roots A_1, A_2; B_1, B_2, etc., are selected, planted in pairs and allowed to cross-fertilize. It is hoped that some of their progeny will possess an

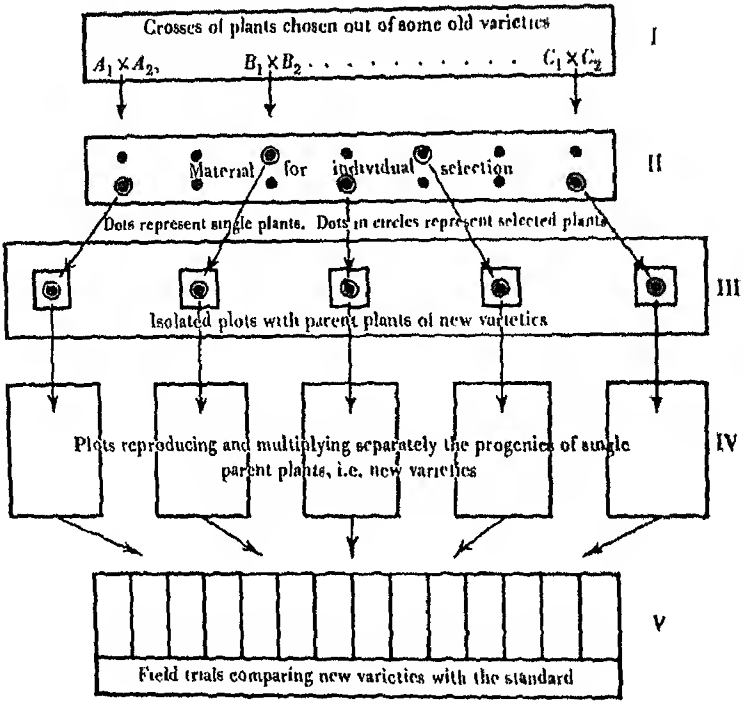


Fig. 1.

increased sugar content. Each of the roots A_1, A_2, etc., being a hybrid with respect to a great number of genes, their progeny will not be homogeneous but will be a mixture of a great number of types of various properties. Therefore a selection from among them is needed. The second step consists in planting the seeds obtained from the crosses on a larger field. In the autumn all roots are lifted and out of each of them a small section is cut out and analysed for sugar content, which does not prevent the root from producing seed if planted again the next season. Roots with small sugar content are discarded and others, promising a sweet progeny, selected for further breeding. This step is called individual selection.

The third step in breeding consists in planting the selected roots in isolated plots, so as to prevent, as far as possible, cross-fertilization. Each of these roots generates a new family of beet and it will be called the parent plant of this variety. The number of seeds it produces is, of course, rather small, and the fourth step, taking up at least two years, consists in multiplying the seeds of the new variety.

The first progeny of the parent plant is sown on a separate plot and allowed to reproduce.

The fifth and final step consists in the test of the results of all preceding steps: all the new varieties are compared in field trials with some established standard. Those which are found to exceed the standard in sugar content are further multiplied and put on the market. The others are discarded.

It is obvious that at all stages described above the breeder is faced with various risks of error.

(1) His choice of roots A_1, A_2; B_1, B_2, etc., used for the cross may be unlucky, and practically all the genetical types produced may have no advantages over the existing standard. This problem, however, lies outside the scope of the present paper.

(2) Even if the cross was a success the breeder may be wrong in his step II and fail to select the proper individuals from which to breed the new varieties. It must be remembered that the individual variation from plant to plant is very large, and it may easily happen that genetically better plants through environmental conditions will be less promising than some of the worse ones. The obvious remedy against overlooking the best genetical types in the process of individual selection is to breed from as many individuals as possible. This is actually often done, but there is a limit to this device imposed by the difficulty in comparing large numbers of new varieties with the standard.

(3) Even if both the cross and the selection of parent plants were successful, the breeder may be unlucky in his field trials. It is known that their accuracy is limited, and it may happen that, through the unavoidable experimental error, the successfully selected new varieties will be judged inferior to the standard, and consequently discarded. In such a case all the previous efforts and expense in breeding and selection of the new varieties would be wasted.

It is obvious that we can avoid this danger by increasing the accuracy of the field trials. Here, however, we come into a conflict with (2). An increased accuracy of field trials means either an increase of the number of replications or an improvement in the method, which, in practice, always means additional expense. If we increase the number of varieties to be compared with the standard, this means another additional expense. So the breeder will ask the question, what is more important, to have more new varieties and test them superficially, or fewer varieties and test them with a great accuracy? This is the problem which will be dealt with in the present paper.

II. THE GENERAL PROBLEM

(a) *Statement of the general problem*

In order to make clear the general problem, consider some particular varieties to be compared with a standard, and denote by X the true excess in sugar content (true excess, for short) which one of these is able to give over the standard

in some particular conditions of soil, treatment and weather. We can never know X, but a field trial may give its estimate, x, which is unavoidably affected by an experimental error. If X is greater than zero, the new variety will be considered as successfully selected. But X may be greater than zero while its estimate, x, owing to the experimental error, may be negative or, even if it is positive, it may be so small that the experimenter will doubt whether its excess over zero does indicate that X is also positive.

If the magnitude of X were known and also the accuracy of the field trial, then it would be possible to calculate the number of replications which are needed to insure that the probability of the trial detecting the fact (in the sense described on p. 34 below) that X is greater than zero, will be as large as desired.

For this purpose it is only necessary to make use of the tables which give the probability of second kind errors, in connection with "Student's" test (see (1) and (2)),* i.e. give the probability that an experiment will fail to detect the advantage in sugar-yielding capacity of the new variety over the standard when it is as large as say, X'. Any seed-breeding station, with an established method of experimentation, can use the results of previous experiments to estimate roughly the standard error per plot to be expected in future, and apply the tables mentioned to calculate how many replications should be made in order that the probability of detecting such varieties which exceed the established standard by any amount X' is as large as, say, 0·8, 0·9, etc. Using this number of replications, the station would feel confident of discovering in 80 or 90 % of trials the varieties exceeding the standard by X'.

This, however, does not solve the problem, because we do not know how frequently the new varieties do exceed the standard by the fixed amount X'. Fixing X' arbitrarily in advance we may fix it so large that the new varieties will practically never give an excess exceeding X' and, thus, the further calculations will be actually useless. In order to obtain useful results, we must know not only how frequently an excess of a given size over the standard will be detected by an experiment of a given accuracy, but also, how frequently excesses of all possible sizes are actually met with in the usual process of breeding and selection. If we know that applying our customary methods we shall usually succeed in selecting varieties which exceed the standard by X_1', X_2', X_3', ... with frequencies, say $P_1, P_2, P_3, \ldots$, we may then apply the tables of the second kind of errors to each of these categories and, thus, see what would be the practical effect of applying any fixed number of replications to the new varieties which are usually presenting themselves for comparison.

It follows that for the solution of the problem of the relation between the number of new varieties and the accuracy of the field trials, the knowledge of the probability that an experiment will detect any specified excess in sugar content is not sufficient. To solve the problem we must also know the distribution of these

* Small figures in brackets refer to literature quoted at the end of the papers.

true excesses over the standard variety which the new varieties may show. It is obviously impossible to make any sure prediction about this distribution, but we may estimate what it has been in the past, and use this estimate to give an idea of what may happen in the future.

The method of estimating the distribution of the true excesses over the standard, shown by a number of varieties in a series of experiments already carried out, is the first problem to be considered.

(b) *The population of the new varieties*

Consider a series of N experiments with the same design and the same number of replications, each comparing the same number, say k, of varieties

$$V_{i1}, V_{i2}, \ldots, V_{ij}, \ldots, V_{ik}, \quad \begin{pmatrix} i=1,2\ldots N \\ j=1,2\ldots k \end{pmatrix} \qquad \ldots\ldots(1)$$

with the same standard variety V_s. Let

$$x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{ik}, \qquad \ldots\ldots(2)$$

be the estimates of the excesses of the varieties (1) over the standard V_s obtained in the ith experiment. It will be noticed that there are Nk different varieties compared with the same standard. Hence, altogether, there will be Nk different x's. It is usually assumed that within any single experiment the standard error, σ_i, of the estimate, x_{ij}, is the same for all varieties. We shall adopt this hypothesis and denote by s_i^2 the unbiased estimate of σ_i^2.

Behind these experimental results, there will be true excesses

$$X_{i1}, X_{i2}, \ldots, X_{ij}, \ldots, X_{ik}, \qquad \ldots\ldots(3)$$

of the varieties compared over the standard. These true excesses (3) depend upon the varieties chosen for trials, which may be regarded as a random sample drawn from a population π described as follows.

Consider first a population, π', of true excesses over the standard which would be observed if *all* the individuals coming from a cross (or, perhaps, crosses) performed by the breeder were used as parent plants of new varieties. Denote by $p'(X)$ the distribution function of the X in that population π'.

Actually, the breeder makes a selection of the individuals from which he intends to breed, and tries to select the best ones. In this, however, he must be sometimes wrong, and we may consider a function $f(X)$ representing the probability that an individual of the population π', capable of generating a variety with the true excess over the standard equal to X will be actually selected by the breeder.

The functions $p'(X)$ and $f(X)$ together determine a certain imaginary population, the one we have denoted by π of the new varieties which, under the

usual conditions of selection, are liable to be compared with the standard. The true distribution of X in this population is, say,

$$p(X) = \text{const.} \times p'(X) f(X), \qquad \text{......(4)}$$

and the varieties which were actually compared with the standard in a particular year may be considered as a random sample from the population π.

It will be noticed that the population π and the distribution $p(X)$ depend on the method by which the parent plants are selected from the population π'. If, for instance, we decide to diminish or to increase the number of the parent plants to be selected, then the distribution $p(X)$ will be changed also. The same will happen if the principle on which the parent plants are selected is altered. It follows that if it be possible to estimate $p(X)$, we may learn something about the suitability of different alternative methods of selecting parent plants.

(c) *The probability P of detecting a "best" variety*

We are now interested in the distribution $p(X)$ of the X's in the population π. Once this distribution is known we can see roughly whether any given size X' of the true excess X, is likely to be met with in practice. Suppose that the true distribution of X's is represented in Fig. 2, where the range of X extends from

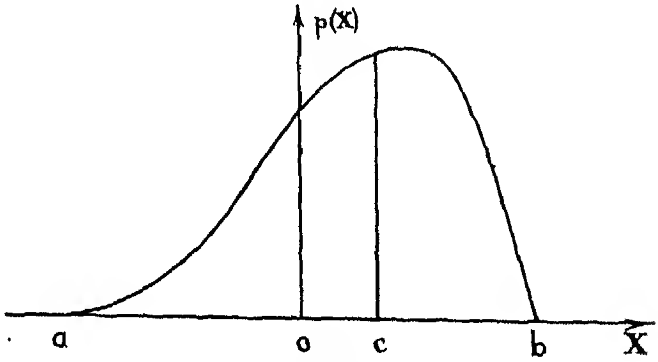


Fig. 2.

a to b. It will be seen that it would be useless to aim in our experiments at the detection of varieties with X exceeding b. In fact, such varieties will never occur. Therefore, the progress in plant breeding depends upon the possibility of identifying those varieties for which X is positive but does not exceed b. In any practical case it will be possible for the breeder to fix a certain value c, lying somewhere between O and b, such that he would consider to be most desirable to detect new varieties with excesses exceeding c. He may, then, adjust the number of replications of his trials so as to have a fair chance of detecting such varieties; for convenience they may be called the "best" varieties. Suppose that we know $p(X)$ and that c is fixed; denote by P the probability that a variety with excess

exceeding c will be detected in a trial of given accuracy. P is easily seen to be given by the formula

$$P = \int_c^b p(X)\, B(X)\, dX, \qquad \text{......(5)}$$

where $B(X)$ is the probability that the field experiment will detect the fact that $X > 0$.

(d) *Kolodziejczyk's results on the power function of "Student's" test and their application to calculate P*

The function $B(X)$, called the power function of the statistical test employed, is easily obtained from the formula given by Kolodziejczyk (3) and we shall discuss it below.

Since each x_{ij} is an estimate of X_{ij}, it is reasonable to assume that x_{ij} is normally distributed about X_{ij} with a standard error σ_i whose estimate, which has been denoted by s_i, is independent of x_{ij}. The joint probability law of x_{ij} and s_i is

$$p(x,s) = \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}(f-2)}\Gamma(\frac{1}{2}f)\,\sigma^{f+1}\sqrt{(2\pi)}}\, s^{f-1}\, e^{-\frac{fs^2+(x-X)^2}{2\sigma^2}}, \qquad \text{......(6)}$$

where f is the number of degrees of freedom used for estimating σ_i.

The statistical method used in analysing the data obtained from field experiments consists in testing the hypothesis H_0, that $X \leqslant 0$, that is to say, that the compared variety is not better than the standard. It has been pointed out by Neyman and Pearson (4) that in testing H_0, two kinds of errors should be considered: the error of rejecting the hypothesis tested when it is true—the first kind of error—and the error of failing to detect that some alternative is true—the second kind of error.

Denote by P_1 the probability of the first kind of error. We may fix in advance any number $0 < \alpha < 1$ which we shall call the level of significance, and arrange the test so that the probability P_1 will never exceed α. For this purpose it is sufficient to make a rule of rejecting the hypothesis tested whenever the ratio $t = x/s$ is greater than t_α, where t_α is the value to be found in R. A. Fisher's tables (5) of t corresponding to $P = 2\alpha$. Below we shall consider two levels of significance, $\alpha = 0{\cdot}05$ and $\alpha = 0{\cdot}01$.

If H_0 is not true and $X' > 0$ is the true value of the excess X, then the chance of the test detecting the fact that X' is greater than zero is evidently

$$B(X') = \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}(f-2)}\Gamma(\frac{1}{2}f)\,\sigma^{f+1}\sqrt{(2\pi)}} \int_0^\infty s^{f-1} e^{-\frac{fs^2}{2\sigma^2}} ds \int_{st_\alpha}^\infty e^{-\frac{(x-X')^2}{2\sigma^2}} dx. \qquad \text{......(7)}$$

Let

$$z_\alpha = \frac{t_\alpha}{\sqrt{f}}, \quad s^2 = \frac{\sigma^2 u^2}{f z_\alpha^2}, \quad z^2 = \frac{(x-X')^2}{\sigma^2}. \qquad \text{......(8)}$$

Substituting these equations into (7) and (5), we get

$$P=\int_c^b p(X)\,B(X)\,dX$$
$$=\frac{1}{\sqrt{(2\pi)}\,2^{\frac{1}{2}(f-2)}\,\Gamma(\frac{1}{2}f)\,z_\alpha^f}\int_c^b p(X)\,dX\int_0^\infty u^{f-1}e^{-u^2/2z_\alpha^2}du\int_{u-x'/\sigma}^\infty e^{-\frac{1}{2}z^2}dz. \qquad \ldots\ldots(9)$$

In order to evaluate the integral (9) accurately, it would be necessary to know the exact nature of the function $p(X)$, and even then the work would probably be rather tedious. Since, however, we cannot hope to know the distribution function $p(X)$ exactly, the best we can do is to get for it a reasonable approximation, using the data of the experiments carried out in previous years. We cannot, therefore, obtain an accurate evaluation of P and shall consider an approximate method of calculating its value given in (9). This will be done by using the exact values of $B(X)$ obtained from the tables referred to(1), and the estimated values of $p(X)$. We shall then apply the simplest quadrature formula,

$$P=\frac{h}{2}(y_0+y_m)+h\sum_{i=1}^{m-1}y_i, \qquad \ldots\ldots(10)$$

where $y_0, y_1, \ldots, y_m$ are the values of the product $p(X)\,B(X)$, calculated at a series of points at equal distance h. The results which it is possible to obtain, using this quadrature formula, will be sufficiently accurate for practical purposes. The approximate information which we may have regarding the function $p(X)$ would not justify the application of any more elaborate method of quadrature.

It is now clear that the knowledge of the function $p(X)$ is essential from the point of view of the problem we are interested in and we shall consider how it could be estimated from the results of previous experiments.

III. Methods of estimating the distribution of X's

(a) *Estimation of $p(X)$ by method of moments*

Write
$$x_{ij}=X_{ij}+\epsilon_{ij}, \quad \begin{pmatrix} i=1,2,\ldots,N \\ j=1,2,\ldots,k\end{pmatrix} \qquad \ldots\ldots(11)$$

where ϵ_{ij} is a random error, which will be assumed to be independent of the X's and normally distributed about zero with the standard deviation σ_i. Denote by $\mathscr{E}(u)$ the expected value of any random variable, u. It is known that

$$\left.\begin{aligned}&\mathscr{E}(\epsilon_{ij}^{2t})=\frac{(2t)!}{2^t t!}\sigma_i^{2t}, \quad (t=1,2,3,\ldots)\\ &\mathscr{E}(\epsilon_{ij}^{2t+1})=0,\end{aligned}\right\} \qquad \ldots\ldots(12)$$

Let m_q' and M_q' be the qth moments of x_{ij} and X_{ij} respectively. From (11) and (12)

$$m_q'=\mathscr{E}(x_{ij}^q)=\mathscr{E}[(X_{ij}+\epsilon_{ij})^q]=\sum_{0\leqslant t\leqslant \frac{1}{2}q}\frac{q!}{2^t t!\,(q-2t)!}M'_{(q-2t)}\sigma_i^{2t}. \qquad \ldots\ldots(13)$$

Putting $q=1, 2, 3$ and 4, we get

$$\left.\begin{aligned} m_1' &= M_1', \\ m_2' &= M_2' + \sigma_i^2, \\ m_3' &= M_3' + 3M_1'\sigma_i^2, \\ m_4' &= M_4' + 6M_2'\sigma_i^2 + 3\sigma_i^4. \end{aligned}\right\} \qquad \text{......(14)}$$

The left-hand sides of (14) represent the moments, about zero, of the observable variate x_{ij} and may be estimated from the experimental data. Solving (14) with regard to M_q' and calculating the central moments M_q of X_{ij} in terms of σ_i^2 and the central moments m_q of x_{ij}, we obtain

$$\left.\begin{aligned} M_1' &= m_1', \\ M_2 &= m_2 - \sigma_i^2, \\ M_3 &= m_3, \\ M_4 &= m_4 - \sigma_i^2(6m_2 - 3\sigma_i^2). \end{aligned}\right\} \qquad \text{......(15)}$$

Strictly speaking equations (15) refer to a particular experiment corresponding to the subscript i. If, however, the accuracy of all experiments is the same and thus $\sigma_1=\sigma_2=\ldots=\sigma_N=\sigma$, then, we can apply the same formulae to all experimental data available. It will be seen below that the assumption that all σ's are equal may be sometimes reasonable and, therefore, we shall adopt here this hypothesis. If it is true that $\sigma_1=\sigma_2=\ldots=\sigma_N=\sigma$ and the number of degrees of freedom for obtaining the common estimate of σ is sufficiently large, we may replace σ^2 by s_0^2, the common unbiased estimate of σ^2. Hence we can estimate the moments of X_{ij} from the observed data in N experiments. Having obtained the moments of X_{ij}, we can calculate $\beta_1(X)$ and $\beta_2(X)$ and determine the corresponding distribution from the Pearson system of curves (6).

Any method of estimation should be tested to see how far it will give reliable results. Especially we want to have some idea as to how accurately we are likely to estimate $p(X)$, using only a limited number of observations. A special theoretical inquiry will be needed to study the efficiency of the method described. Until such work is completed it was thought useful to test the method empirically and two artificial examples were worked out. However, before proceeding to these examples, I shall describe an alternative method of estimating $p(X)$ due to Eddington and described by Levy and Roth (7).

(b) *Alternative method of estimating $p(X)$*

At this stage it will be convenient to alter a little the notation concerning the probability laws. Denote by $u(x)$ the probability law of x, the estimate of X; $p(X)$, as before, will represent the probability law of the true excesses; ϵ will denote the difference between x and X and $p(X, \epsilon)$ the simultaneous probability

law of X and ϵ. We have assumed that the experimental error ϵ is independent of X and normally distributed about zero with standard deviation σ, the value of which is assumed to be possible to estimate accurately from a large number of experiments. It follows that

$$p(X, \epsilon) = p(x)\frac{1}{\sqrt{(2\pi)}\,\sigma} e^{-\epsilon^2/2\sigma^2}, \qquad \ldots\ldots(16)$$

where σ may be considered as known.

Introduce now a new system of variables

$$\left.\begin{aligned} X &= x + \eta, \\ \epsilon &= -\eta, \end{aligned}\right\} \qquad \ldots\ldots(17)$$

the simultaneous probability law of x and η will be found as

$$p(x, \eta) = \frac{1}{\sqrt{(2\pi)}\,\sigma} p(x+\eta)\, e^{-\eta^2/2\sigma^2} \qquad \ldots\ldots(18)$$

In order to obtain the probability law of x, we have to integrate this expression with regard to η

$$u(x) = \frac{1}{\sqrt{(2\pi)}\,\sigma}\int_{-\infty}^{\infty} p(x+\eta)\, e^{-\eta^2/2\sigma^2}\, d\eta. \qquad \ldots\ldots(19)$$

This corresponds exactly to the first formula in Levy's book (7), p. 157. On the next page he gives an expansion which makes it possible to calculate the value of $p(X)$ in terms of the values of $u(x)$ and its successive differences, viz.

$$p(X) = u(x) - \frac{\sigma^2}{2}\Delta^2 u(x) + \frac{\sigma^2}{2}\Delta^3 u(x) - \frac{\sigma^2}{2}\left(\frac{11}{12} - \frac{\sigma^2}{4}\right)\Delta^4 u(x) + \ldots \qquad \ldots(20)$$

This method has been tried as an alternative in the examples discussed below.

(c) *Empirical test of the two methods of estimating $p(X)$*

Both examples which we shall describe below consist in assuming arbitrary distributions $p(x)$ and in obtaining by laboratory methods a set of figures which could be obtained as experimental data if the assumed hypothesis and the distributions $p(X)$ were in fact true. I started in the two cases with the assumption that the true distributions, $p(X)$, were represented by the histograms shown in Figs. 3 and 4. In order to obtain the x's, it was necessary to add to each value of X a random error ϵ, independent of X and normally distributed about zero. These were obtained from the tables of normal deviates published by Mahalanobis (8). These deviates represent what would be the observed values of a random variable, ϵ, normally distributed about zero with its standard deviation equal to unity.

Adding normal deviates to the values of X, I have obtained 100 numbers and these were then considered as the values of the x's and were used to estimate $p(X)$

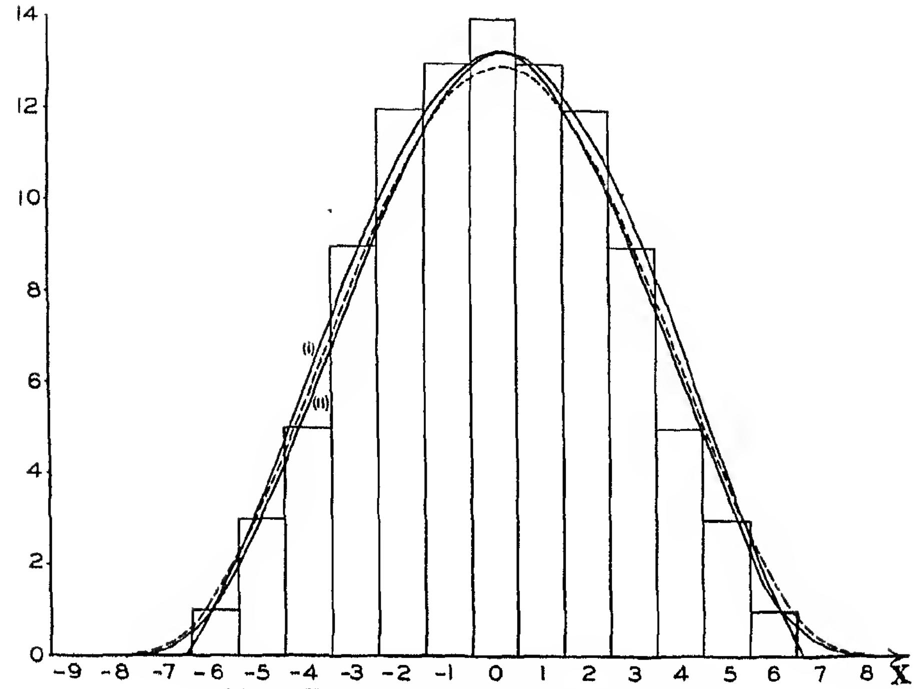


Fig. 3. True distribution of X and its estimates.

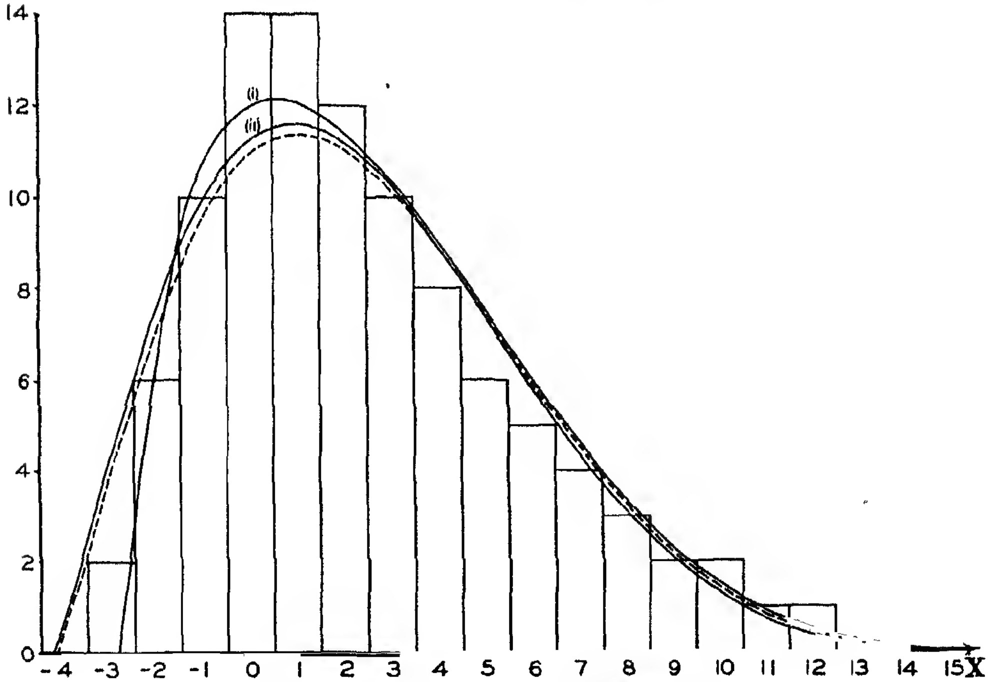


Fig. 4. True distribution of X and its estimates.

- - - - - Fitted distribution of x.

———— (i) Fitted distribution of X (by method of moments).

———— (ii) Fitted distribution of X (by Levy's formula).

by the two methods described. These methods, however, need also the estimate of σ^2. In order to have a situation analogous to that which we have in practice, I performed another random sampling experiment, and obtained 20 values of s^2 by sampling from the known distribution of s^2 with a fixed value of $\sigma^2 = 1$ and with the number of degrees of freedom equal to 25. The arithmetic mean of those s^2's was used as a common estimate of σ^2. The same method would be applied in practice to the data of a series of 20 experiments of equal accuracy, each comparing five new varieties with a standard in six randomized blocks. Having applied the methods described to the results of the sampling experiments, the estimates of $p(X)$ were obtained which may be compared with the true distribution from which we started.

In order to obtain the set of values of s^2, it is necessary to apply the usual sampling technique with Tippett's random numbers (9) and the distribution of s^2

$$p(s^2) = \frac{f^{\frac{1}{2}f}}{2^{\frac{1}{2}f}\,\Gamma(\frac{1}{2}f)\,\sigma^f}(s^2)^{\frac{1}{2}f-1}e^{-fs^2/2\sigma^2}, \qquad \text{......(21)}$$

The distribution, $p(X)$, in Example 1 was assumed to be symmetrical about zero, while in Example 2 it was asymmetrical. The frequencies are given in Table I.

TABLE I

Hypothetical distribution of X

X	−6	−5	−4	−3	−2	−1	0	1	2	3	4	5	6	7	8	9	10	11	12
1st example	1	3	5	9	12	13	14	13	12	9	5	3	1						
2nd example				2	6	10	14	14	12	10	8	6	5	4	3	2	2	1	1

Table II gives the values of frequency constants obtained from the observed values of x's, using for $p(X)$ the formulae (15).

TABLE II

Frequency constants for $p(x)$ and $p(X)$

	$p(x)$			$p(X)$	
	Example 1	Example 2		Example 1	Example 2
m_1'	0·1405	2·5850	M_1'	0·1405	2·5850
m_2	7·8233	11·4415	M_2	6·8773	10·3255
m_3	0·4027	19·4065	M_3	0·4027	19·4065
m_4	148·6389	367·9831	M_4	106·9186	295·1070
$\beta_1(x)$	0·0003	0·2564	$\beta_1(X)$	0·0005	0·3421
$\beta_2(x)$	2·4191	2·8110	$\beta_2(X)$	2·2606	2·7680

The Pearson Curves fitting $p(x)$ and $p(X)$, found by the method of moments in the two cases, are as follows:

Example 1:

$$p(x) = 12{\cdot}9321\left(1 - \frac{x^2}{65{\cdot}1587}\right)^{2{\cdot}6644}, \qquad \ldots\ldots(22)$$

$$p(X) = 13{\cdot}2480\left(1 - \frac{X^2}{42{\cdot}0525}\right)^{1{\cdot}5573}, \qquad \ldots\ldots(23)$$

both curves are with origin at their common mean 0·1405.

Example 2:

$$p(x) = 11{\cdot}3920\left(1 + \frac{x}{5{\cdot}3513}\right)^{1{\cdot}4771}\left(1 - \frac{x}{16{\cdot}8592}\right)^{4{\cdot}6536}, \qquad \ldots\ldots(24)$$

$$p(X) = 12{\cdot}1229\left(1 + \frac{X}{3{\cdot}5190}\right)^{0{\cdot}7231}\left(1 - \frac{X}{14{\cdot}6820}\right)^{3{\cdot}0168} \qquad \ldots\ldots(25)$$

The origins of the two curves are at their respective modes, 1·0486 and 0·6400.

In Figs. 3 and 4 the histograms represent the true distributions of X, the dashed curves correspond to equations (22) and (24) the curves marked (i) to equations (23) and (25), while the curves marked (ii) represent the estimates of $p(X)$ obtained by Levy's method. It is seen that in both cases, both methods of estimating $p(X)$ give satisfactory results.

Of course, the sampling experiments cannot be considered as a definite evidence that a particular method of estimation is satisfactory, however favourable may be the results. However, the two examples described above seem to be encouraging, and we may hope that the results obtained below by applying our method to the data of actual experiments give us reasonable approximations to the true distributions of X.

(*d*) *The case where σ varies from experiment to experiment*

In the above theory we have made an essential assumption that the standard error of the estimated excesses of the new varieties over the standard does not change from one experiment to another. This is a possible hypothesis in case where all the experiments considered are carried out on a single large field by the same experimenter with the same care. However, we must be clear, as far as possible, whether this hypothesis is justified or not. First we may test it by the usual L_1-test (10). If this gives a favourable result then the application of the above method may be considered as more or less safe. But the L_1-test may provide evidence that the standard error σ does vary from experiment to experiment. This, however, is not necessarily sufficient to make the above methods of estimating $p(X)$ totally invalid. In fact, the variation of σ may exist, but it may be within a very small range, and in this case we may expect that the result of the

estimation of p (X), based on the assumption that σ is throughout constant, will not be very inaccurate.

Example 3. It would be difficult to study theoretically what inaccuracy in any particular case may arise in estimating $p(X)$ by the method of moments described above, when σ is not constant. In order, however, to throw some light on this point another sampling experiment, similar to those in Examples 1 and 2, was carried out. It was assumed that in each of the 20 hypothetical experiments in which X varied as in Example 2 the values of σ were different. The distribution of σ was assumed to be as given in the following table.

TABLE III

Hypothetical distribution of σ

σ	0·65	0·70	0·75	0·80	0·85	0·90	0·95	1·00	1·05	1·10	1·15	1·20	1·25	1·30	1·35
Frequency	1	1	2	1	1	1	2	2	2	1	1	1	2	1	1

Mean $\sigma=1$; S.D. of $\sigma=0{\cdot}2$

This is a relatively wide spread so that the example should provide a fairly severe test of the adequacy of the method based on the assumption that σ is constant. Out of this distribution, using Tippett's random numbers, a random sample of 20 was drawn and the values of σ's obtained were associated with 20 hypothetical experiments. To obtain the errors involved in the x's, the original values of ϵ's, which were obtained in the previous sampling, were multiplied by the corresponding values of σ. This would exactly correspond to random sampling from a normal population with the particular value of σ.

The variation in the true values of σ from one experiment to another would also affect the estimate of the variance of x, the change being proportional to σ^2. Accordingly, the values of s^2 obtained previously for each hypothetical experiment were multiplied by the appropriate σ^2.

Having thus obtained a new set of empirical data of 20 hypothetical experiments with varying σ, the previous method of moments was applied to estimate $p(X)$, and the results are shown in Fig. 5.

The histogram, as previously, represents the true distribution of X; the continuous curve represents its estimate obtained previously when σ was constant from experiment to experiment; and lastly, the dashed curve represents the estimated $p(X)$ obtained by the same method from data affected by the variation of σ. It is seen that the two curves differ, but not very seriously. This may be considered as an indication that when the variation of σ from experiment to experiment is only moderate, our method may still be used to provide a reasonably accurate estimate of $p(X)$. This fact is important, because even if the L_1-test fails to detect the variation in σ this may still exist.

IV. APPLICATION TO THE ACTUAL EXPERIMENTAL DATA

In the following, I apply the method described to experimental data with sugar beet which were kindly supplied by Messrs K. Buszczyński and Sons, Ltd., Warsaw, and it is a pleasure to express here my gratitude to the Directors of this firm. The data used refer to the experiments carried out in 1923 and 1924, in

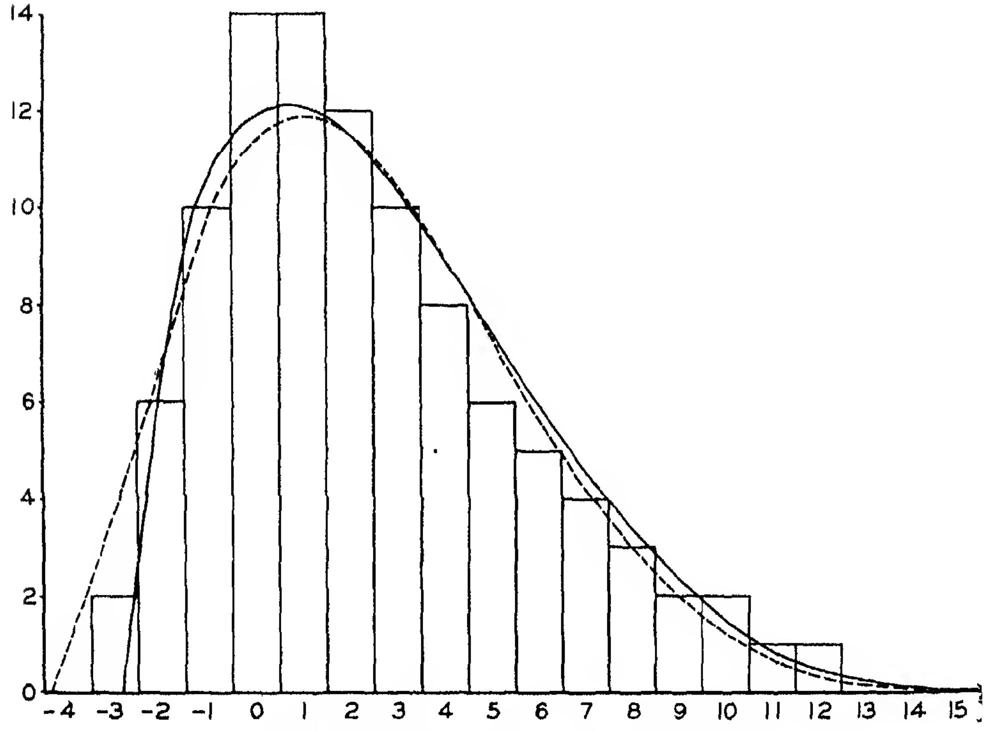


Fig. 5. Effect of variability of σ on efficiency of method of estimating $p(X)$.

———— Estimated distribution of X's when σ's are equal.

- - - - - - Estimated distribution of X's when σ's are different and the coefficient of variation of σ is 20% of mean σ. Histogram is the true distribution of X's.

one of the firm's experimental stations, Górka Narodowa. The total number of experiments carried out each year was about 100, each comparing with the standard three new varieties selected and bred by the firm. All these experiments were carried out on a very large and uniform field by the same staff and using the same methods. This circumstance makes it probable that the assumption of the standard error in each experiment being constant, or at least not very variable, is not far from being true. The number of replications was not the same in all experiments. In order to get the material into a form convenient for numerical work, i.e. to have the same number of replications in each experiment, out of each

year's data 40 experiments were selected, each with 5 replications. About the layout of these experiments I have the following information: the experimental plots were comparatively narrow and long, and cut across the direction of ploughing so as to make them as homogeneous as possible. The number of roots in each plot was 100. Of course, during the vegetative period some of them perished. The distribution of the varieties in each particular experiment was systematic, as shown in Fig. 4.

V_s	V_{t1}	V_{t2}	V_{t3}	V_s	V_{t1}	V_{t2}	V_{t3}	V_s	V_{t1}	V_{t2}	V_{t3}	V_s	V_{t1}	V_{t2}	V_{t3}	V_s	V_{t1}	V_{t2}	V_{t3}	V_s
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10

Fig. 6. Arrangement of experiments.

The systematic arrangement of the experiments did not permit the use of the customary methods of working out the data, as those assume randomization. The method used was that proposed by Neyman(10), consisting in estimating the fertility level for each plot and each variety. The basic assumptions being that: (i) a fourth order parabola is able to represent the level with sufficient accuracy, and (ii) that the levels corresponding to two different varieties are parallel.

It will be noticed that with the systematic arrangement as shown in Fig. 6, the comparison between varieties V_{t1} and V_s, V_{t3} and V_s must be more accurate than that of V_{t2} and V_s. In all cases we have the same number of replications and in the former case the difference in soil in adjoining plots sown with the compared varieties must be, on the whole, smaller than in the other. This intuitive inference is numerically expressed in Neyman's formulae and in the final results, but the estimated variances of the excesses of V_{t1} over V_s, and of V_{t2} over V_s appeared to be very close, e.g. $s_{11}^2 = s_{13}^2 = 0{\cdot}0117$, $s_{12}^2 = 0{\cdot}0118$. For this reason these differences were ignored.

Tables IV and V show the values of the x's and s^2's calculated for each of the varieties compared in years 1923 and 1924. The L_1-test applied to the 40 s^2's did not discover any significant variation in their size, in 1923. It was found in fact that $L_1 = 0{\cdot}910$ whereas $L_1\,(0{\cdot}05) = 0{\cdot}906$ for $f = 13$, $N = 40$. On the other hand the variation in s^2 in 1924 proved to be significant, $L_1 = 0{\cdot}854$. It follows that while the data for 1923 gave us no reason to doubt the validity of the assumption that σ is constant, it is possible that the variation in σ in 1924 will influence unfavourably the accuracy of the estimate of $p\,(X)$.

Here, however, we may remember the encouraging results of the sampling experiment discussed above as Example 3, which shows that the method of moments is not very sensitive to moderate variation in σ. Of course, it would be desirable to carry out this experiment assuming that the distribution of σ is approximately what it actually was in the experiments considered. For that purpose an attempt was made to estimate this distribution on the lines similar to

No. of exp. ...	1	2	3	4	5	6	7	8	9	10
x_1 x_2 x_3	0·4569 0·2328 0·2884	0·9867 0·8133 0·7603	0·2130 0·2950 0·7382	0·5460 0·2686 0·5826	0·5877 0·6183 0·2485	0·4000 0·2242 0 4543	0 5587 0·7386 0·7058	0·4908 0·5296 0·4396	0·4293 0·2640 0·4365	0·6542 0·2202 0·7506
s^2	0·0117	0·0211	0·0135	0·0199	0·0084	0·0227	0·0236	0·0171	0·0210	0·0133
No. of exp. ...	11	12	13	14	15	16	17	18	19	20
x_1 x_2 x_3	0·3713 0·3300 0·2363	0·4357 0·5966 0·4716	0·4252 0·4638 0·3901	0·6427 0·4795 0·2714	0·3554 0·0952 0 3577	0·4239 0·5510 0·1206	0·4199 0·3901 0·0450	0·1338 −0·0195 −0·0377	0·0559 −0·0653 −0·0221	0·3141 0·3685 0 5951
s^2	0·0103	0·0082	0 0098	0·0308	0·0124	0·0085	0·0078	0·0364	0·0118	0 0099
No. of exp. ...	21	22	23	24	25	26	27	28	29	30
x_1 x_2 x_3	0·5446 0 4334 0·5698	0·4907 0·0207 0·2334	0·2166 −0·0598 −0·3514	0 3592 0·0634 0·1096	0·4954 0·2854 0·0933	0·2528 0·2304 0·0589	0·2605 −0 7483 0·2166	−0 1432 −0·1938 −0·1595	−0·0094 −0·1581 −0·1287	−0·2431 0·6335 0·4402
s^2	0·0074	0·0132	0·0087	0·0158	0 0242	0·0141	0 0369	0·0154	0·0086	0·0214
No. of exp. ...	31	32	33	34	35	36	37	38	39	40
x_1 x_2 x_3	0 2011 0·3882 0·4141	0·3853 0 4198 0·2993	0·1465 −0·0882 −0·0808	0·4023 0·5970 0·9688	0·6617 0·4167 0·5848	0·5422 0·2967 0·0801	−0·1517 −0 5452 −0·1074	0·3591 0·2851 0·3177	−0·0832 0 1093 0·2749	0·2849 0·4060 0·4886
s^2	0·0175	0 0202	0 0104	0·0164	0 0078	0 0243	0·0155	0·0143	0·0150	0·0140

TABLE V

Values of x (as per cent. of sugar content in beet) and s^2 obtained in 40 experiments, 1924

No. of exp. ...	1	2	3	4	5	6	7	8	9	10
x_1	0·1378	0·0069	0·1829	−0·1209	−0·2120	−0·6658	−0·6658	0·2031	0·1911	0·1337
x_2	0·3911	−1·0381	0·2675	−0·1197	0·2402	−0·3308	0·1864	0·2582	−0·7977	0·1062
x_3	−0·4613	0·2335	0·5709	−0·2183	−0·2258	−1·1006	−1·0178	−0·3380	0·2146	0·3000
s^2	0·0327	0·0540	0·0469	0·0138	0·0570	0·0203	0·0222	0·0207	0·0078	0·0342
No. of exp. ...	11	12	13	14	15	16	17	18	19	20
x_1	0·0103	−0·0248	−0·1146	0·5159	0·4788	0·3330	−0·3323	−0·1815	0·1209	0·1989
x_2	−0·0521	0·1289	0·1536	0·2945	−0·0128	0·2304	−0·3291	−0·1627	−0·0339	0·3626
x_3	0·0308	0·3215	0·2700	0·7708	−0·0687	0·3457	0·1742	−0·4800	0·1126	0·3775
s^2	0·0096	0·0112	0·0296	0·0368	0·0201	0·0317	0·0282	0·0767	0·0279	0·0145
No. of exp. ...	21	22	23	24	25	26	27	28	29	30
x_1	−0·0391	−0·5041	0·3331	0·3351	−0·3050	−0·6337	−0·0664	0·4107	−0·3837	−0·1021
x_2	−0·4864	−0·9305	0·3992	0·1939	−0·2696	−0·3598	0·0577	−0·1181	−0·6595	−0·1928
x_3	0·3081	−0·5072	0·3849	−0·0694	−0·1928	−0·2844	−0·0860	−0·3603	−0·7246	−0·0374
s^2	0·0095	0·0533	0·0138	0·0158	0·0218	0·0395	0·0160	0·0271	0·0296	0·0134
No. of exp. ...	31	32	33	34	35	36	37	38	39	40
x_1	−0·4058	−0·4709	−0·3317	0·2347	−0·1932	0·2024	−0·3552	−0·2669	0·4315	0·3110
x_2	−0·2326	0·0059	−0·2188	0·0308	0·0021	0·2767	−0·1759	−0·3137	0·3182	0·2788
x_3	−0·5930	−0·1101	−0·5986	−0·4354	−0·6183	0·1150	−0·4649	−0·3280	0·0045	0·1988
s^2	0·0132	0·0156	0·0161	0·0332	0·0308	0·0071	0·0189	0·0293	0·0272	0·0108

those followed to estimate $p(X)$. However, a few sampling experiments carried out to test the method, showed that its efficiency is very poor. Consequently it was abandoned. On the other hand, the estimate Σ^2 of the variance of the σ's based on that of the observed s_i's, which the reader will have no difficulty in calculating, namely

$$\Sigma^2 = s_0^2 - (\bar{s})^2 \tfrac{1}{2} f \left(\frac{\Gamma(\frac{1}{2}f)}{\Gamma\left(\frac{f+1}{2}\right)} \right)^2, \qquad \text{......(26)}$$

where s_0^2 means, as formerly, the arithmetic mean of the observed variances s_i^2 and $\bar{s}$ that of their square roots s_i, proved to be fairly accurate. This formula was applied to the experimental data of 1924 and it was found that Σ amounted to about 20 % of $\bar{s}$. This empirical result was used to fix the variation of σ in the sampling experiment of Example 3, so as to have its S.D. also equal to 20 % of the mean. The results obtained there suggest that applying the method of moments to estimate $p(X)$ for the experimental data of 1924, we should not be very wrong.

The usual frequency constants calculated for the two years for the distributions of x and X are as follows:

1923	1924
$s_0^2 = 0.0160$	$s_0^2 = 0.0259$
$m_2(x) = 0.0832$	$m_2(x) = 0.1358$
$\beta_1(x) = 0.3490$	$\beta_1(x) = 0.1683$
$\beta_2(x) = 3.9609$	$\beta_2(x) = 2.8962$
$\beta_1(X) = 0.6617$	$\beta_1(X) = 0.3179$
$\beta_2(X) = 4.4718$	$\beta_2(X) = 2.8412$

Here $m_2(x)$ denotes the variance of x. It will be noticed that the ratio m_2/s_0^2 is in the two cases just over 5, which is of about the same order as in the sampling experiments carried out to test the efficiency of estimating $p(X)$.*

The values of the β's for both $p(x)$ and $p(X)$ suggested type V and type I Pearson Curves in the years 1923 and 1924, respectively. These lead to the following equations for $p(x)$ and $p(X)$ obtained using the method of moments:

$$1923:\quad p(x) = 4.72334\,(10)^{38}(-x)^{-50.8241}\, e^{\frac{97.4138}{x}}, \qquad \text{......(27)}$$

origin at $x = 2.2970$;

$$p(X) = 4.27389\,(10)^{17}(-X)^{-29.1419}\, e^{\frac{35.9874}{X}}, \qquad \text{......(28)}$$

origin at $X = 1.6277$;

* For the sampling experiments the values of $m_2(x)$ are given in Table II, p. 40 above, while σ^2 was unity.

$$1924: \quad p(x) = 1\cdot064\left(1+\frac{x}{2\cdot3073}\right)^{0\cdot1518}\left(1-\frac{x}{0\cdot8569}\right)^{3\cdot3989}, \qquad \ldots\ldots(29)$$

origin at $x = 0\cdot0260$;

$$p(X) = 1\cdot1738\left(1+\frac{X}{1\cdot6359}\right)^{4\cdot0770}\left(1-\frac{X}{0\cdot4464}\right)^{1\cdot1124} \qquad \ldots\ldots(30)$$

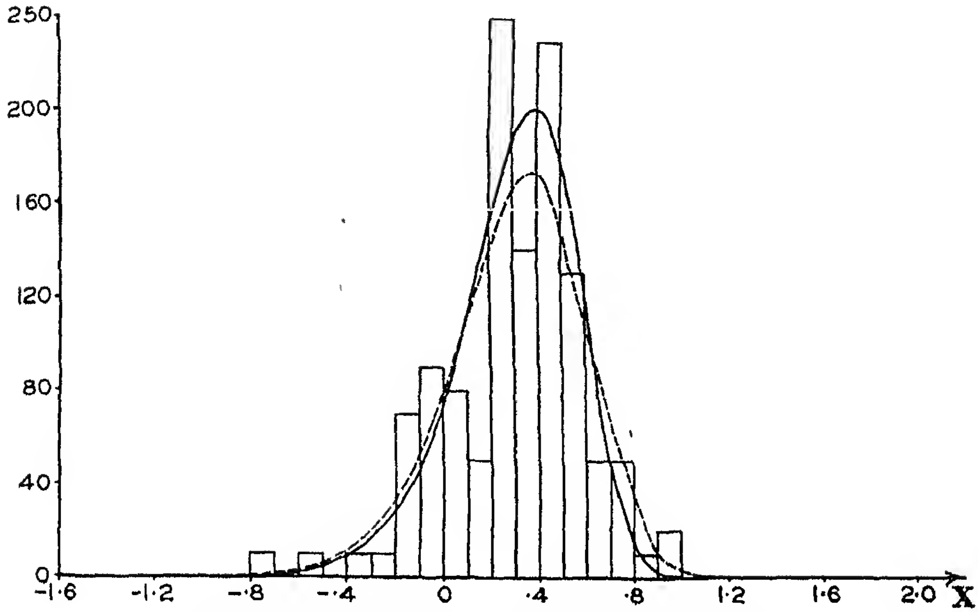
origin at $X = 0\cdot0918$.

The curves are represented in Figs. 7 and 8, where the histograms refer to the observed values of x's, the continuous curves represent the estimated $p(X)$ and the dashed curves represent $p(x)$.

It is seen that in the two years the curves differ both in shape and relative position with respect to the origin of co-ordinates. This may be due partly to the change in atmospheric conditions and partly to the fact that the standard variety used was not the same in the two years. In fact, the variety which was used as a standard was the one which, in the previous year's competitive experiments carried out by a special commission appointed by the sugar industry in Poland, proved to be the sweetest. This change in the standard varieties is probably justified in the special conditions of sugar beet breeding. However, in other cases as, for instance, in breeding of barleys for brewing, the standard variety would be probably more stable.

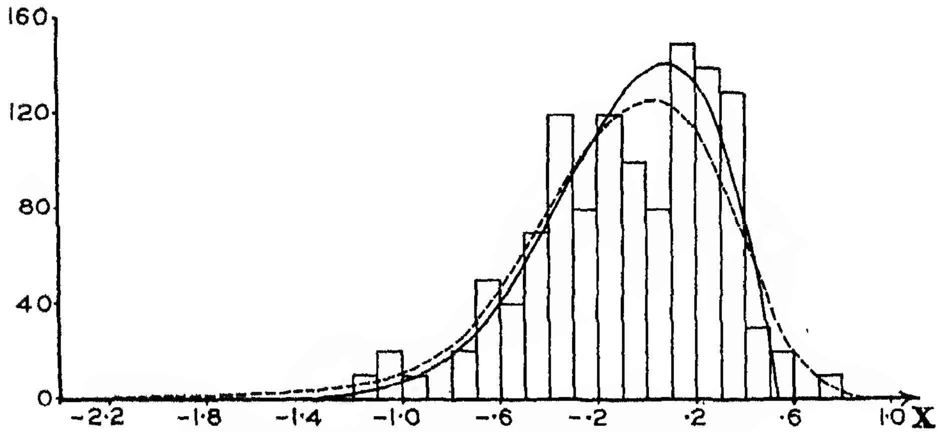
Having this in view, we shall have to consider two possible ways of proceeding: one corresponds to the assumption that the standard variety remains unchanged from year to year, and the other to the case where the standard variety is changed. Because of lack of experimental data corresponding to the first situation, we shall explain the procedure, using the material concerning the sugar beet described above and ignoring the circumstance that the standard variety was in fact not the same in the two years. Thus, the shift in the curve will be ascribed solely to the changes in atmospheric conditions. In order to illustrate the second situation we shall use the same data, taking into account the fact that the standard variety was different in the two years.

We shall now start by considering the first situation. Let us agree to call "good" varieties, in each year, those which proved to be sweeter than the standard. The percentages of these could be found by calculating the areas under the curves, $p(X)$, extending to the right-hand side of the origin of coordinates. The calculations showed that, in the year 1923, there were about 87·5 % of good varieties and, in the year 1924, about 46·5 %. 50 % of the sweetest out of the good varieties may be called the "best" varieties. In the year 1923, the best varieties will be those exceeding the standard by 0·37 % of sugar content, and in 1924 this limit will be 0·23 %. Now, we may calculate the probability of detecting a good or a best variety in an experiment with some particular number of replications, if the accuracy of those experiments were equal to that of the actual experiments.



Excess of sugar content as a percentage of standard

Fig. 7. Estimated distribution of X from actual experiments, 1923



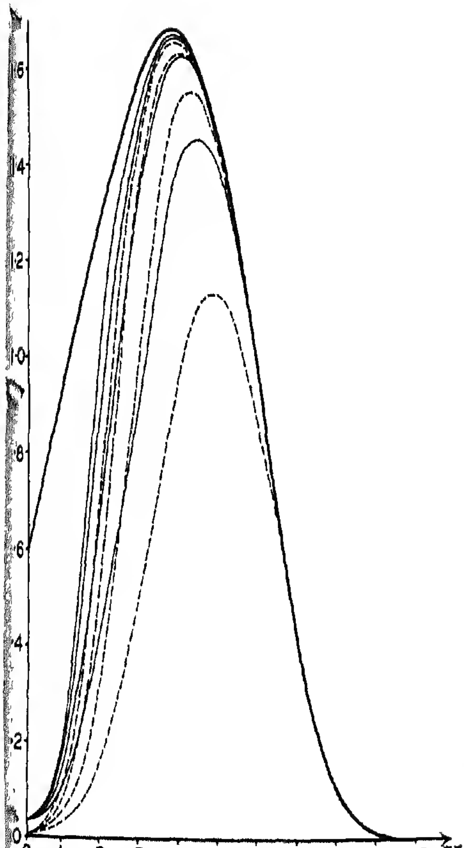
Excess of sugar content as a percentage of standard

Fig. 8. Estimated distribution of X from actual experiments, 1924

Histogram: Observed values of x.
- - - - - Distribution of $p(x)$.
———— Distribution of $p(X)$.

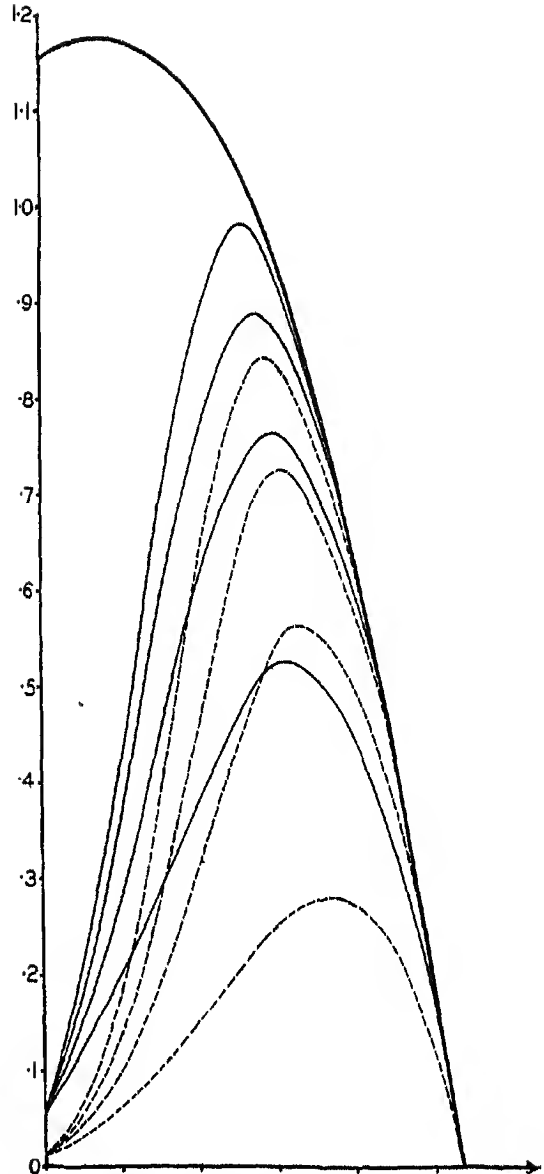
Applying the method described above (section II, p. 36), Figs. 9 and 10 were constructed. The outer thick curve represents the part of the distribution $p(X)$ taken from Figs. 7 and 8 extending to the right of the origin of coordinates. The area under this curve is, in each case, equal to unity, which means that we limit our consideration to the good varieties only. The ordinates of all other curves were obtained by multiplying $p(X)$ by the corresponding values of the power function $B(x)$ obtained from the Neyman tables. Areas under the continuous and the dashed curves represent the probabilities of detecting a good variety with the sugar excess falling within any given limits, if the number of replications were $n=5$, 10, 15 and 20. It was assumed here that the accuracy of those hypothetical experiments, that is to say the standard error per plot, is equal to that of the actual ones, but that the layout of the experiments is different, namely, they were assumed to be arranged in randomized blocks. The difference between the continuous and the dashed curves is that the former correspond to the case where the assumed level of significance (the probability of first kind errors) is $\alpha=0{\cdot}05$ and the latter when it is $\alpha=0{\cdot}01$. It is seen, as could be expected, that, in the latter case, the detecting power of the experiments is considerably smaller. Tables VI and VII give the probabilities of detecting any of the good and any of the best varieties in accordance with the number of replications and with the level of the significance used. These probabilities are areas under the continuous and the dashed curves in Figs. 9 and 10. For the best varieties these areas had to be doubled.

Figs. 9 and 10 and Tables VI and VII may be used to draw conclusions as to the number of replications to be used in the following years. Our attention must be directed primarily towards the best varieties. Looking at the tables we see that the conditions in the two successive years differ enormously: while in 1923, five replications make the chance of detecting a best variety, at $\alpha=0{\cdot}05$, equal to 0·908, the same chance in 1924 was only 0·542. This is due to the change in the standard error per plot connected with weather conditions, and also to the shift of the curve with respect to the origin. In 1923 the standard error per plot was 0·199 and in 1924 it increased to 0·254. Rational planning of future experiments requires obviously the knowledge of changes in accuracy of the experiment occurring from year to year. Two years' observations indicate only that the variation may be very great. According to the prevailing possibilities of using space, additional labour, etc., when planning experiments for the third year we may take into account the possibilities of weather conditions giving as low an accuracy of experiments as in 1924. Then it might be thought advisable to use as many as 10 or 15 replications. If, however, such a scale of experiments is for various reasons prohibitive, it may be necessary to use a smaller number of replications. Looking at Table VII, we see that if $n=5$, and if the accuracy of the experiments is as bad as in 1924, then it would not be wise to apply the level of significance $\alpha=0{\cdot}05$, let alone $\alpha=0{\cdot}01$.



Excess of sugar content as a percentage of standard

Fig. 9. Probabilities of detecting good varieties in conditions of 1923



Excess of sugar content as a percentage of standard

Fig. 10. Probabilities of detecting good varieties in conditions of 1924

Distribution of true sugar excess:

———— in population of varieties tested.

———— in population of varieties likely to be found significant at 0·05.

- - - - - in population of varieties likely to be found significant at 0·01.

N.B. The four curves of each type, starting from the highest, relate to cases $n = 20$, 15, 10 and 5 respectively

The procedure to be advised in this case seems to be as follows. If the economic conditions have forced the breeder in some future year to use only 5 replications, the decision as to what varieties should be considered as failing to exceed the standard, should be based on the analysis of the whole lot of the experiments as given in the present paper. If the calculations lead to figures as in Table VII, for 1923, then it would mean that the accuracy of the experiments was satisfactory and probably there would be no objection to the use of the level of significance

TABLE VI

Chance of detecting a "good" variety

n	1923		1924	
	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$
5	0·649	0·483	0·343	0·166
10	0·756	0·665	0·485	0·319
15	0·815	0·744	0·558	0·410
20	0·841	0·781	0·609	0·478

TABLE VII

Chance of detecting a "best" variety

n	1923		1924	
	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$
5	0·908	0·756	0·542	0·288
10	0·974	0·953	0·707	0·559
15	0·996	0·991	0·865	0·711
20	0·999	0·994	0·922	0·814

$\alpha=0{\cdot}05$ or even $0{\cdot}01$. In fact, the application of $\alpha=0{\cdot}05$ will detect over 90 % of all best varieties and nearly 65 % of all good varieties. The remainder may be neglected. If, however, the calculations lead to a picture similar to what we found for 1924, then it would be advisable to apply special precaution in order not to discard the varieties the value of which may be considerable. The best thing to do would be to classify the varieties tested into the following groups: (i) those for which the advantage over the standard was proved beyond any reasonable doubt, even under the prevailing unfavourable conditions; (ii) those for which the value of x is not significant at $\alpha=0{\cdot}01$ but is so at some greater values of α, perhaps at $\alpha=0{\cdot}1$ or more, this value being so chosen that the probability of

detecting a "best" variety is considerable, say 0·9 or more;* (iii) the third group will consist of the remaining varieties which it will be more or less safe to discard. Obviously, it is difficult to give any general rule discriminating between what is to be considered as a large and a small chance of detecting a best variety. This must be left to persons responsible for the whole experimental work and the process of breeding. The problem of the statistician is accomplished when he finds means of calculating this chance. Of course, if the number of replications is very considerable, then all these calculations may not be necessary. But this probably will be only rarely the case.

TABLE VIII

Chance of detecting a "good" variety

n	1923		1924	
	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$
5	0·903	0·729	0·319	0·155
10	0·969	0·941	0·452	0·297
15	0·992	0·980	0·519	0·382
20	0·996	0·991	0·567	0 445

TABLE IX

Chance of detecting a "best" variety

n	1923		1924	
	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$	$\alpha=0{\cdot}05$	$\alpha=0{\cdot}01$
5	0·976	0·891	0·518	0·269
10	0·994	0·979	0·739	0·526
15	0·999	0·997	0·842	0·675
20	1·000	0·999	0·904	0·780

Finally, we must consider the situation which presents itself when the standard variety is changing from year to year. In this case, the experimenter will have to consider two points: first, if the distribution $p(X)$ is situated almost entirely to the left of the origin of coordinates, this may indicate that his method of breeding

* I may remark that to carry out these calculations Neyman's tables of probabilities of second kind errors should be extended so as to apply to other levels of significance beyond the $\alpha=0{\cdot}05$ and $\alpha=0{\cdot}01$.

and selecting is not satisfactory. The error may lie in the choice of the parent plants used for crosses. Again, there may be something wrong in his principle of selection of single plants generating new families. This point lies beyond the limit of the present paper. Secondly, the experimenter will be interested in the possibility of making a proper choice out of the existing material. He will probably use another definition of the best and good varieties.* For example, he may define the best varieties to be those the sugar content of which exceeds that of 75 % of the whole material. Again, the good varieties may be defined as those which exceed in sugar content, say, 50 % of the whole lot. It is very easy to calculate the tables analogous to Tables VI and VII corresponding to these new definitions. The results are given in Tables VIII and IX. The discussion is quite similar to that given above.

V. Summary of results

The whole process of plant breeding may be roughly divided into two parts: (i) the production of new families or varieties which may prove to be better than the established standards, and (ii) the test whether any of these new varieties do exceed in quality the established standards. The second of these steps is connected with field trials in which the new varieties are compared with the variety taken as a standard.

The quality of any variety is a very complex conception and depends on a large number of different characters. However, there is usually some single character of the plants, the importance of which is greater than that of any others, and which by itself is being taken as a conventional measure of the quality. This may be the average yield, the sugar content, percentage of nitrogen, etc. The difference between the average value of such a character in a new variety and in the standard is called an excess over the standard, which may be either positive or negative. The field trials are not able to give the true values of the excesses but only their estimates which are necessarily affected by experimental errors. Through these experimental errors it is possible that the new varieties with positive and perhaps even relatively considerable excess will not be detected, which may lead to their ultimate rejection. It is obvious that such a circumstance is unsatisfactory as it involves considerable waste of effort connected with a successful breeding of a new variety. The question therefore arises as to what number of replications in field trials should be used in order to have a fair chance of detecting new varieties with positive and sufficiently large excesses over the standard. The solution of this problem requires the knowledge, or at least an approximate knowledge, of the distribution of the true excesses (not of their estimates) over the standard, likely to be found in new varieties which may present themselves for comparison with the standard. Of course, this distribution

* See p. 48 above.

is connected with the method of breeding. A method of obtaining an estimate of the distribution of the true excesses, based on the examination of the results of similar trials in previous years, is the main topic of the present paper. The method devised was tested on a few artificially constructed sampling experiments, then compared with an alternative method advanced by Eddington and Levy, and found to be satisfactory. It was then applied to actual experimental data concerning 120 new varieties of sugar beets bred for sugar content by Messrs K. Buszczyński and Sons, Ltd., Warsaw, and tested in the same conditions on two adjoining fields in the years 1923 and 1924. Having obtained the estimate of the frequency distribution of the true excesses in each year, it was then possible to judge the efficiency of future experiments with $n = 5, 10, 15$ and 20 replications, in detecting the "good" and "best" varieties, if the accuracy of the experiments were similar to those in 1923 and 1924.

Some general conclusions as to the number of replications to be used and as to the method of procedure if the accuracy of the fields proves to be poor have been drawn. The method of estimating the distribution of true excesses may be useful also when two different methods of selecting new varieties are compared. And here we come to the original question formulated at the beginning of the paper: which course is better, to start say 200 new varieties each year and then test them with 5 replications only, or to diminish the number of new varieties to some 100 and test them with 10 replications? If the records of sugar content of parent plants of 200 varieties already in field trials are available, then the breeder is able to see what would have been his results if he had started only 100 of them. Using all the 200 varieties, he would be in the position to estimate, say $p_{200}(X)$, the distribution of true excesses among these 200 new varieties and also, say $200 \times P(200, 5)$, the number of best varieties which he may reasonably expect to detect in trials with 5 replications. Again, he may use the records of the sugar content (and probably of other properties) of the parent plants to see what would be the results of his individual selection if he had decided to start only 100 new varieties. Picking out of the records of the field trials the data concerning the 100 varieties which would have been selected in such a case, he would be in the position to estimate, say $p_{100}(X)$, the distribution of the true sugar content among these varieties. This distribution would lead him to, say $100 \times P(100, 10)$, the expected number of best varieties which would be detected in the field trials with 10 replications. The comparison between $100 \times P(100, 10)$ and $200 \times P(200, 5)$ would provide the answer to the question formulated above.

In conclusion, I wish to express my hearty thanks to Dr J. Neyman for suggesting this problem to me and for his constant help, both during the course of research and whilst writing the paper.

REFERENCES

(1) Neyman, J. (1935). "Statistical problems in agriculture experimentation." *Supp. J. R. Statist. Soc.* **2**, 131–2.

(2) Neyman, J. and Tokarska, B. (1936). "Errors of the second kind in testing 'Student's' hypothesis." *J. Amer. Statist. Ass.* **31**, 318–26.

(3) Kołodziejczyk, St. (1933). "Sur l'erreur de la seconde catégorie dans le problème de M. Student." *C.R. Acad. Sci., Paris*, **197**, 814.

(4) Neyman, J. and Pearson, E. S. (1936). "Contributions to the theory of testing statistical hypotheses." *Stat. Res. Mem.* **1**, 1–5.

(5) Fisher, R. A. (1934). *Statistical Methods for Research Workers*, 5th ed., Table IV. Oliver and Boyd.

(6) Pearson, K. (1930). *Tables for Statisticians and Biometricians*, Part I, 66. *Biometrika* Office.

(7) Levy, H. and Roth, L. (1936). *Elements of Probability*. Oxford: Clarendon Press.

(8) Mahalanobis, P. C. (1934). "Tables of random samples from a normal population." *Indian J. Statist.* **1**, 303–28

(9) Tippett, L. H. C. (1927). Random Sampling Numbers. *Tracts for Computers*, No. 15. Camb. Univ. Press.

(10) Nayer, P. P. N. (1936). "An investigation into the application of Neyman and Pearson's L_1-test, with tables of percentage limits." *Stat. Res. Mem.* **1**, 38–51.

(11) Neyman, J. (1929). "The theoretical basis of different methods of testing cereals. Part II." *Wiadomości Mat. Warsaw.*

RURAL MORTALITY. ITS COMPARATIVE SEX INCIDENCE

By W. J. MARTIN and E. A. CHEESEMAN

Of the Medical Research Council's Statistical Staff

From the Division of Epidemiology and Vital Statistics, London School of Hygiene and Tropical Medicine

The mortality of England and Wales has exhibited certain characteristics, of which the two most prominent have been the subject of various investigations. The first of these is geographical; proceeding from the north southwards a decrease in mortality is observed. The second is the excess of urban mortality over rural. In addition, attention has been drawn to a further feature which would also appear to be of a permanent nature, i.e. the ratio of the death-rate in rural districts to that of the general population is proportionately lower for males than for females. The purpose of this short paper is to attempt as far as is possible from the available data an inquiry into this phenomenon.

The data used were the recorded deaths for males and females during the triennial periods 1920–2 and 1930–2 in the aggregated rural districts of the whole country and its geographical divisions. The ratio of actual to expected deaths in each area was calculated. Utilizing the records for 1920–2 as an example, the expected deaths were obtained as follows. The average male death-rates during that period for England and Wales, in single years for the first five years of life, in quinquennial groups from ages 5 to 85 and of one group of age 85 and over were applied to the rural male population, at corresponding ages, recorded at the 1921 census for each of the areas indicated in Table I. The sums of the resultant series gave the numbers of expected male deaths for each area. A similar procedure was adopted for the females. The actual deaths were taken as the average of the triennia and the index tabulated was that of the actual deaths divided by the expected. The results are given in Table I.

Table I shows that the males enjoyed a relatively more favourable mortality than the females for both the triennia 1920–2 and 1930–2. For all the areas combined the male deaths were 19 and 15 % below the number expected on the basis of the whole country for 1920–2 and 1930–2 respectively, while the rural female deaths were 12 and 8 % less. In each area the ratio was larger for females than for males in each triennial period. The differences between the male and female ratios were small in the South-east, South-west and Midlands I divisions. The largest differences existed in the two Welsh and in the Northern I division, where in both triennia the actual female deaths were in excess of the expected.

A large difference was also noted in 1930–2 for the Northern III division, where once again the actual female deaths were greater than the expected.

In view of the disparity in the ratios of the two sexes at all ages, the analysis was next made for specific age periods. The results are shown in Table II. From this table it appeared that the relatively more favourable male mortality was not confined to any particular age period, but was evident in every age group above age 15 in every division of the country. Under age 15 there was little or no difference between the male and female ratios for the country as a whole, but within some divisions some variation occurred.

TABLE I

Deaths from All Causes in Rural Districts

Actual deaths/Expected deaths

Area	1920–2		1930–2	
	Males	Females	Males	Females
South-east	0·72	0·76	0·78	0·81
North I	1·02	1·17	1·00	1·18
North II	0·81	0·88	0·84	0·96
North III	0·98	1·04	0·96	1·09
North IV	0·86	0·94	0·88	0·96
Midlands I	0·83	0·85	0·86	0·91
Midlands II	0·81	0·90	0·86	0·94
East	0·73	0·82	0·77	0·87
South-west	0·77	0 83	0·84	0·88
Wales I	0·94	1·06	1·02	1·13
Wales II	0 93	1·05	0·97	1·09
England and Wales	0·81	0·88	0·85	0·92

To determine if any particular cause of death was responsible for the observed differences the deaths of the whole country were divided into seven broad categories. Expected deaths were calculated as before, using the death-rates at ages from each category for England and Wales as a whole. The ratios of actual to expected deaths were tabulated and shown in Table III for all ages and for four age groups over 15. For all ages, every cause of death with the exception of violence showed a relatively greater decrease among males than among females. The greatest differences between male and female ratios were those for pulmonary tuberculosis and cancer, where the males showed a relatively greater improvement than did the females.

The excess of the female ratios over the males from these two causes of death was common to every division of the country, as can be observed from Table IV. Generally a large difference between the male and female ratios of actual to

TABLE II

Deaths from All Causes in Rural Districts classified according to Age and Area

Actual deaths/Expected deaths

Age group	0–5		5–15		15–25		25–35		35–45		45–55		55–65		65–75		75 and over	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Area	1920–2																	
South-east	0·57	0·55	0·71	0·67	0·71	0·82	0·78	0·83	0·71	0·78	0·66	0·73	0·68	0·73	0·73	0·77	0·88	0·87
North I	1·21	1·23	1·08	1·18	1·06	1·05	0·93	1·31	0·85	1·20	0·83	1·05	0·92	1·14	0·96	1·12	1·03	1·15
North II	0·77	0·71	0·78	0·80	0·83	0·93	0·88	0·98	0·68	0·86	0·67	0·83	0·72	0·90	0·83	0·90	0·96	0·99
North III	1·17	1·11	1·03	0·97	0·90	0·95	0·82	1·14	0·85	0·96	0·74	0·99	0·90	0·98	0·97	1·03	0·99	1·04
North IV	0·73	0·75	0·93	0·80	0·86	0·79	0·89	0·94	0·85	0·88	0·78	0·93	0·85	0·92	0·92	1·08	1·02	1·08
Midlands I	0·70	0·71	0·77	0·77	0·86	1·03	0·85	0·93	0·77	0·90	0·76	0·80	0·78	0·79	0·88	0·87	0·96	0·93
Midlands II	0·79	0·77	0·89	0·90	0·75	1·02	0·78	1·00	0·73	0·94	0·69	0·90	0·75	0·88	0·81	0·92	0·95	0·97
East	0·66	0·64	0·71	0·75	0·82	1·14	0·88	0·98	0·66	0·85	0·61	0·82	0·63	0·75	0·71	0·78	0·88	0·91
South-west	0·59	0·56	0·70	0·67	0·74	0·99	0·83	0·92	0·76	0·86	0·72	0·85	0·75	0·80	0·80	0·85	0·90	0·92
Wales I	0·94	0·95	0·96	1·00	1·10	1·36	0·96	1·27	0·89	1·15	0·89	1·06	0·89	1·10	0·93	1·05	0·96	1·01
Wales II	0·84	0·85	1·12	1·05	1·02	1·37	1·03	1·24	0·81	1·19	0·86	1·07	0·89	1·00	0·96	1·04	1·04	1·09
England and Wales	0·76	0·75	0·83	0·82	0·84	1·00	0·85	0·99	0·76	0·91	0·72	0·86	0·76	0·84	0·82	0·88	0·93	0·94
	1930–2																	
South-east	0·63	0·62	0·81	0·73	0·84	0·76	0·81	0·83	0·79	0·77	0·69	0·84	0·74	0·80	0·78	0·80	0·88	0·87
North I	1·19	1·22	1·10	1·22	1·21	1·30	1·17	1·15	0·94	1·22	0·81	1·04	0·88	1·17	0·94	1·15	1·04	1·20
North II	0·82	0·82	0·85	0·77	0·79	0·92	0·95	0·85	0·70	0·95	0·73	0·88	0·78	0·92	0·83	1·00	0·95	1·06
North III	1·19	1·17	1·27	1·43	1·06	1·11	0·88	1·05	0·87	1·21	0·81	0·97	0·85	1·02	0·91	1·04	1·00	1·08
North IV	0·85	0·79	0·81	0·84	0·87	0·66	0·79	0·94	0·72	0·94	0·78	0·91	0·87	0·97	0·91	1·05	1·00	1·04
Midlands I	0·79	0·80	0·75	0·83	0·87	0·97	0·92	0·96	0·83	0·99	0·78	0·85	0·84	0·88	0·85	0·88	0·97	0·97
Midlands II	0·88	0·91	0·87	0·77	0·93	0·94	0·93	0·89	0·82	0·98	0·75	0·91	0·78	0·93	0·86	0·95	0·96	0·99
East	0·72	0·72	0·79	0·72	0·83	0·96	0·78	1·02	0·71	0·91	0·65	0·91	0·69	0·84	0·76	0·82	0·88	0·92
South-west	0·72	0·71	0·80	0·66	0·87	0·93	0·90	1·00	0·81	0·92	0·77	0·89	0·83	0·87	0·83	0·87	0·91	0·93
Wales I	1·00	1·05	0·95	1·10	1·15	1·31	1·17	1·25	0·99	1·15	1·02	1·17	0·99	1·10	1·02	1·16	1·02	1·10
Wales II	0·95	0·94	1·01	1·13	1·10	1·40	1·11	1·39	0·99	1·20	0·88	1·08	0·93	1·06	0·96	1·05	0·99	1·10
England and Wales	0·82	0·82	0·88	0·86	0·92	0·95	0·90	0·98	0·82	0·95	0·76	0·91	0·81	0·91	0·84	0·91	0·93	0·96

expected deaths from pulmonary tuberculosis was associated with a large difference between the ratios from cancer. Not only was this association evident within the divisions but it was also exhibited in each triennial period, that is to say the divisions where the female ratio showed a large excess over the male in 1920–2 also generally showed a large difference in 1930–2.

TABLE III

Deaths in Rural Districts classified according to Age and Cause of Death

Actual deaths/Expected deaths

	1920–2									
	All ages		15–		25–		45–		65–	
	M	F	M	F	M	F	M	F	M	F
Influenza	0·88	0·90	0·88	1·08	0·80	0·92	0·79	0·82	0·98	0·93
Pulmonary tuberculosis	0·69	0·91	0·80	0·96	0·76	0·94	0·57	0·85	0·59	0·90
Other respiratory diseases	0·64	0·67	0·67	0·87	0·56	0·75	0·54	0·57	0·69	0·70
Cancer	0·84	0 93	0·84	1·03	0·83	0·91	0·77	0·88	0·91	0·96
Circulatory	0·87	0 92	0·67	0·73	0·81	0·75	0·79	0·88	0·92	0·96
Violence	0·96	0·85	1·06	1·10	1·04	0·92	0·99	0·84	0·82	0·75
Other causes	0·86	0·92	0·86	1·12	0·86	1·05	0·80	0·90	0·92	0·98

	1930–2									
	All ages		15–		25–		45–		65–	
	M	F	M	F	M	F	M	F	M	F
Influenza	1·06	1·10	0·98	1·01	1 04	1·12	0·97	1·03	1·11	1·13
Pulmonary tuberculosis	0·65	0·88	0·68	0·85	0·73	0·94	0·56	0·85	0·63	0·89
Other respiratory diseases	0·73	0·77	0·71	0·82	0·70	0·85	0·64	0·70	0·79	0·80
Cancer	0·86	0·95	0·80	1·04	0·82	0·88	0·81	0·93	0·91	0·97
Circulatory	0·83	0 89	0 70	0·70	0·73	0·78	0·75	0·88	0·86	0·91
Violence	1·05	0·88	1·31	1·33	1·19	0·98	1·05	0·85	0·78	0·73
Other causes	0·90	0·97	0·95	1·11	0·90	1·08	0·85	0·96	0·94	1·02

It has long been known that occupation affects mortality, and the opinion has also been expressed that migration is an additional factor contributing to the high comparative mortality in young adult life in the rural areas. It has been observed that females migrate from the rural districts at an earlier age than males. If, as has often been suggested, only the healthier persons move to the towns then the excessive female migration would have an adverse affect on the rural

TABLE IV

Deaths in Rural Districts for All Ages classified according to Area and Cause of Death

Actual deaths/Expected deaths

Area	Influenza		Pulmonary tuberculosis		Other respiratory diseases		Cancer		Circulatory		Violence		Other causes	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F
1920–2														
South-east	0·75	0·75	0·67	0·77	0·49	0·54	0·88	0·90	0·83	0·83	0·79	0·76	0·75	0·78
North I	1·48	1·65	0·73	1·08	1·04	1·14	0·77	0·91	0·87	1·05	1·29	0·86	1·08	1·24
North II	0·85	0·99	0·59	0·89	0·54	0·62	0·87	0·95	0·92	0·94	1·01	0·86	0·88	0·93
North III	1·02	1·08	0·65	0·77	0·98	0·96	0·84	0·98	0·92	1·01	1·23	1·10	1·04	1·11
North IV	0·83	0·91	0·59	0·66	0·76	0·83	0·84	0·90	0·90	1·03	1·01	1·03	0·93	1·01
Midlands I	1·01	0·99	0·67	0·87	0·67	0·65	0·80	0·88	0·94	0·93	0·99	0·90	0·85	0·88
Midlands II	0·76	0·78	0·55	0·85	0·69	0·72	0·85	0·98	0·80	0·91	0·97	0·87	0·87	0·97
East	0·72	0·70	0·70	1·07	0·48	0·53	0·85	0·96	0·76	0·81	0·81	0·81	0·80	0·88
South-west	0·73	0·78	0·73	0·92	0·53	0·85	0·81	0·89	0·89	0·92	0·89	0·80	0·81	0·87
Wales I	1·21	1·31	0·83	1·16	0·89	0·89	0·77	0·88	0·86	1·09	1·37	0·87	0·95	1·13
Wales II	0·95	1·01	1·04	1·39	0·73	0·79	0·95	1·06	1·06	1·14	0·89	0·84	0·96	1·07
England and Wales	0·88	0·90	0·69	0·91	0·64	0·67	0·84	0·93	0·87	0·92	0·96	0·85	0·86	0·92
1930–2														
South-east	0·88	0·91	0·65	0·74	0·62	0·67	0·85	0·92	0·77	0·80	1·02	0·89	0·79	0·82
North I	1·21	1·17	0·77	1·15	1·00	1·19	0·81	1·04	0·94	1·12	1·15	0·94	1·09	1·26
North II	1·01	1·18	0·51	0·79	0·69	0·70	0·84	0·94	0·86	1·02	1·06	0·91	0·89	1·01
North III	1·08	1·16	0·65	0·76	0·99	1·02	0·83	1·08	0·94	1·05	1·14	0·97	1·03	1·18
North IV	0·96	1·05	0·54	0·61	0·72	0·83	0·86	0·96	0·93	1·05	1·10	1·01	0·92	0·98
Midlands I	1·18	1·22	0·58	0·86	0·74	0·75	0·86	0·91	0·85	0·90	1·14	1·01	0·90	0·94
Midlands II	1·18	1·15	0·60	0·81	0·80	0·84	0·85	0·94	0·80	0·87	1·03	0·86	0·94	1·05
East	0·97	1·01	0·59	0·98	0·60	0·73	0·86	0·94	0·73	0·80	0·89	0·76	0·84	0·93
South-west	1·18	1·20	0·67	0·88	0·65	0·66	0·87	0·90	0·82	0·85	0·95	0·79	0·90	0·97
Wales I	1·38	1·49	0·81	1·27	1·02	1·04	0·89	0·99	0·98	1·08	1·31	0·97	1·06	1·20
Wales II	1·12	1·20	0·97	1·45	0·80	0·84	0·95	1·10	0·91	1·06	1·06	0·70	1·05	1·17
England and Wales	1·06	1·10	0·65	0·88	0·73	0·77	0·86	0·95	0·83	0·89	1·05	0·88	0·90	0·97

female mortality, This factor has been offered as an explanation of the relatively high rural female mortality in the age group 15–25.

Thus two further features emerge which might be put forward as contributing to the phenomenon under consideration. An examination of this aspect must of necessity be limited owing to the lack of suitable data. An attempt was made, however, to divide the rural areas of the country according to these two headings, (1) occupation, and (2) migration. No unit smaller than a county could be taken. The counties were distributed in relation to the occupations followed by the male inhabitants of their rural areas and were ultimately classified into three broad groups, (1) those with less than 33 %, (2) those with 33 to 50 % and, (3) those with more than 50 %, of the rural males engaged in agriculture.

Turning next to the problem of migration, the only measure readily accessible was a ratio of female to male inhabitants of the rural areas for each county. For England and Wales as a whole the sex ratio was, in 1921, 109·5 females per 100 males and in 1931 the ratio was 108·8. To indicate the extent of female migration, the counties have been divided into three categories of ascending sex ratio, (1) less than 100 %, (2) 100–104 % and (3) more than 104 %.

Table V was drawn up to show the ratio of actual to expected deaths in accordance with these two groupings. For each of the occupational groups the female ratio was in excess of the male for both triennia. The non-agricultural group had the largest ratio and differed significantly from the other two groups, whilst there was practically no difference between the ratio of the second and third occupational groups. The female ratios exceeded those of the males by an almost constant amount throughout and did not indicate that occupation was a cause of the discrepancy between the male and female ratio of actual to expected deaths.

The group with the lowest sex ratio had the highest ratio of actual to expected deaths. This group differed significantly from each of the remaining two, between which there was no difference. The excess of the female ratio over the male steadily declined with increasing sex ratio. The difference between the first and third group, in both triennia, was statistically significant.

The standard error of the ratios, r, given in this table was taken to be $\sigma_r = \sqrt{D}/E$, where D = actual deaths and E = expected deaths. This result was arrived at as follows:

Notation: P' = population in the age groups considered, d' = death-rate per unit in this group; P and d are the corresponding quantities in the same age group in England and Wales; Σ denotes summation over all age groups and $V(u)$ = sampling variance of any quantity, u.

If the actual deaths in any group only differ through chance from the England and Wales value, then

$$D = \Sigma(P'd'), \quad V(D) = \Sigma\{P'^2 V(d')\},$$

where

$$V(d') = \frac{d(1-d)}{P'}. \qquad \ldots\ldots(1)$$

TABLE V

Deaths from All Causes in Rural Districts classified according to Occupation and Sex Ratio

Actual deaths/Expected deaths

	1920–2		
	Males	Females	Differences
Percentage of occupied males engaged in agriculture			
(1) Less than 33 %	0·89 ± 0·0063	0·97 ± 0·0069	0·08 ± 0 0093
(2) 33 to 50 %	0·75 ± 0·0056	0·81 ± 0·0061	0·06 ± 0·0083
(3) Over 50 %	0·76 ± 0·0083	0·84 ± 0 0092	0·08 ± 0·0124
Differences between (1) and (2)	0·14 ± 0·0084	0·16 ± 0·0092	—
(1) and (3)	0·13 ± 0·0104	0·13 ± 0·0115	—
(2) and (3)	0·01 ± 0·0100	0·03 ± 0·0110	—
Females/Males			
(1) Less than 100	0·90 ± 0·0077	0·99 ± 0·0088	0·09 ± 0·0117
(2) 100–104	0·77 ± 0·0059	0·83 ± 0·0065	0·06 ± 0·0088
(3) Over 104	0·79 ± 0·0063	0·84 ± 0·0066	0·05 ± 0·0091
Differences between (1) and (2)	0·13 ± 0·0097	0·16 ± 0·0109	—
(1) and (3)	0·11 ± 0·0099	0·15 ± 0·0110	—
(2) and (3)	0·02 ± 0·0086	0·01 ± 0·0093	—

	1930–2		
	Males	Females	Differences
Percentage of occupied males engaged in agriculture			
(1) Less than 33 %	0·92 ± 0·0065	1·00 ± 0·0073	0·08 ± 0·0098
(2) 33 to 50 %	0·81 ± 0·0059	0·86 ± 0·0063	0·05 ± 0·0086
(3) Over 50 %	0·80 ± 0·0086	0·89 ± 0·0098	0·09 ± 0·0130
Differences between (1) and (2)	0·11 ± 0·0088	0 14 ± 0·0096	—
(1) and (3)	0·12 ± 0·0108	0·11 ± 0·0122	—
(2) and (3)	0·01 ± 0·0104	0·03 ± 0·0117	—
Females/Males			
(1) Less than 100	0·93 ± 0·0080	1·04 ± 0·0094	0·11 ± 0·0123
(2) 100–104	0·81 ± 0·0061	0·88 ± 0·0068	0·07 ± 0·0091
(3) Over 104	0·84 ± 0·0065	0·89 ± 0·0069	0·05 ± 0·0095
Differences between (1) and (2)	0·12 ± 0·0101	0·16 ± 0·0116	—
(1) and (3)	0·09 ± 0·0103	0·15 ± 0·0117	—
(2) and (3)	0·03 ± 0·0089	0 01 ± 0·0097	—

N.B. The figures after the ± sign are standard errors.

Hence approximately $$V(D) = \Sigma(P'd) = E. \qquad \ldots\ldots(2)$$

If the hypothesis of chance variation were true, then we should substitute the true death-rates into (1). In the present case, however, we know from the general consistency of the results and from previous knowledge that the death-rates are lower in the rural areas; it seemed therefore better to take

$$V(d') = \frac{d'(1-d')}{P'} \qquad \ldots\ldots(3)$$

giving approximately $$V(D) = \Sigma(P'd') = D \qquad \ldots\ldots(4)$$

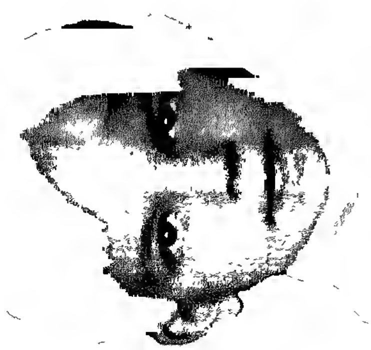
or $$\sigma_D = \sqrt{D}.$$

It will be legitimate to neglect the error in E compared with the error in D, owing to the larger populations on which the death-rates in England and Wales are based; we find therefore $$\sigma_r = \sqrt{D}/E. \qquad \ldots\ldots(5)$$

Conclusions

In relation to the general death-rate in England and Wales, the mortality in the rural areas during the triennia 1920–2 and 1930–2 was proportionately lower for males than females. This was not only true of these aggregated areas, but also of the rural areas within the major divisions of the country and was apparent at all ages above age 15. The causes of death which largely contributed to this were, as far as can be judged, phthisis and cancer. Emigration may be an influential factor. The sex ratio (male/female) of the populations in rural areas favoured the male. If this fact is accepted as evidence of a greater exodus of female migrants there arises the probability that the residual female population is the more unhealthy. This suggestion seemed to be confirmed by the fact that where the sex ratio in the population was highest, the divergence in the ratios of the actual to expected deaths for males and females was lowest.

Teddy P.'s father (IV 3)



Teddy P. (V 1)

A PIEBALD FAMILY

By A. M. NUSSEY

PIEBALDS are sufficiently rare to warrant the publication of a hitherto unrecorded family.

Cockayne (1933) gives a full account of this condition in his book under the heading of "Abnormalities of Pattern", and a bibliography will be found there. The condition behaves as a dominant and three types are described.

My piebald family would fall into the subgroup with a white frontal blaze and pigmented dorsal stripe, the remaining two varieties being one with no frontal blaze and white dorsal stripe, and the other with no white frontal blaze and dorsal surface pigmented.

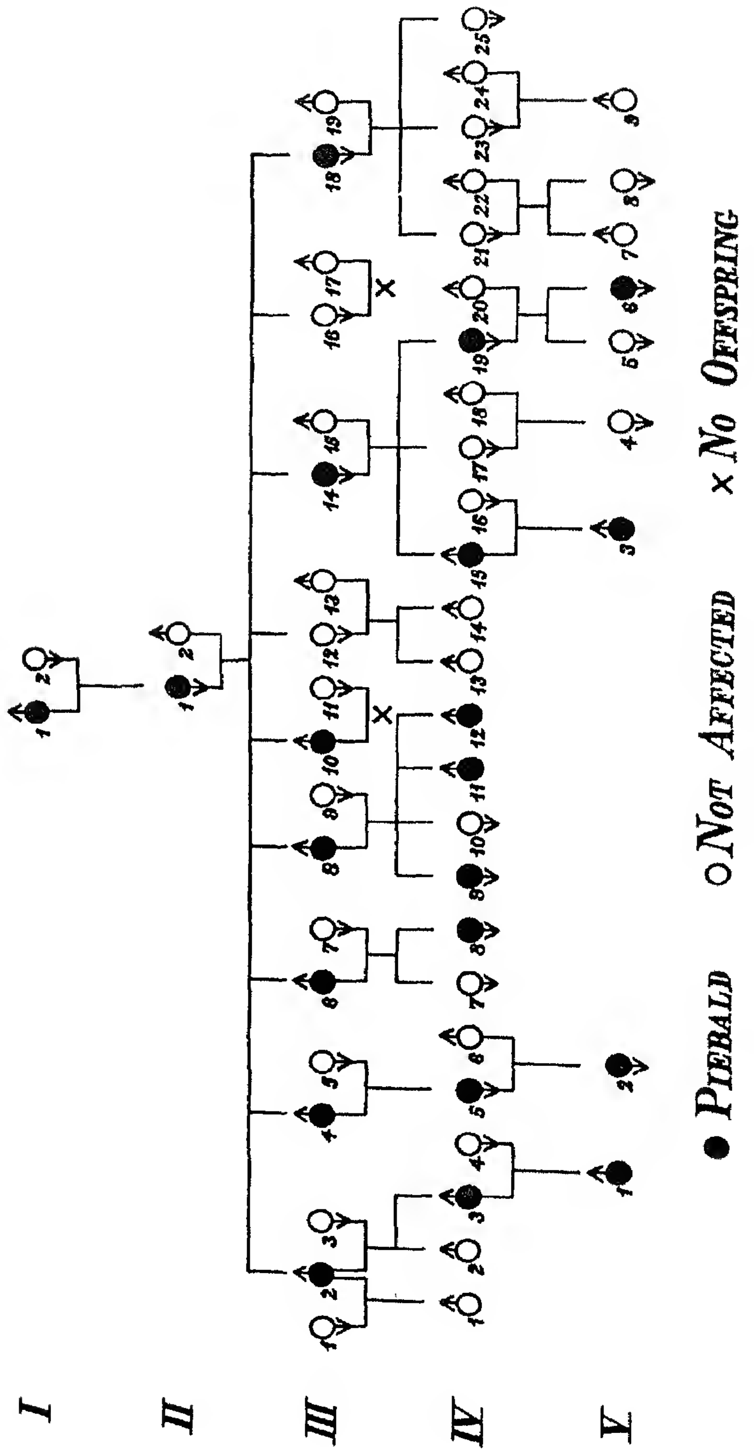
Only three piebald families have been recorded before in England: the London one by Bishop Harman (1909), and two other families described by Cockayne (1914, 1935) both of which came from Suffolk. My piebald family is domiciled in and around Birmingham, and as far as I could find did not originate in Suffolk. Tradition has it that the first piebald in the family was a Frenchman who settled in England about 100 years ago.

Unfortunately the family which I am about to describe showed great reluctance to come forward, and as a result this article is not fully documented.

The only individual whom I was privileged to see and photograph is Teddy P. (V 1). He is of dark complexion, has dark eyes (no heterochromia) and shows a frontal blaze, unpigmented patch of skin in the centre of the forehead (only faintly visible in the photograph), a small white patch to the right of the umbilicus, an extensive patch in the front of the upper part of the left leg, and another but smaller patch in the corresponding position of the right leg.

The boy's mother assured me that the father (IV 3) and all the other affected members of the family show exactly the same markings. The father is very sensitive about his white forelock to the extent of keeping his cap on almost continuously, but I was able to obtain a photograph of him as a youth showing the white blaze. I subsequently saw the father (IV 3) and confirmed that the distribution of the white blaze and other unpigmented patches is practically identical with that of his son. The boy's grandmother (III 3) agreed with the details of the tree, and told me that other affected members were also very sensitive about the white blaze. An uncle (III 4) went so far as to apply chemicals to disguise it and as a result lost his hair in that region.

I 1 (C.) is said to have exhibited the typical markings and to have transmitted it to II 1 (P. *née* C.), but nothing is known about their sibs.



Pedigree of family.

II 1 had nine children, five boys (III 2, 4, 6, 8, 10), all of whom were affected, and four girls, two of whom (III 12, 16) escaped.

Subsequent transmission occurs, as is usual with dominant characters, only through affected members whether male or female, and so we see that III 2 who married twice and had three boys (IV 1, 2, 3) transmitted it to one of them (IV 3) and through him to V 1. The same happens in the case of III 4, IV 5 and V 2; in III 6 and IV 8; III 8 and IV 9, 11, 12; III 14, IV 15 and V 3; and III 14, IV 19 and V 6.

The total number of members in the five generations is 38, of which 21 were affected and 17 were not. It must, however, be taken into account that six of the latter (IV 13, IV 14, V 4, 7, 8 and 9) could not, according to Mendelian laws, have exhibited the abnormality, and so if one deducts also the first two piebalds (I 1, II 1) about whose sibs nothing is known, this gives a proportion of 19 piebalds to 11 who did not exhibit this character, which is close enough to the expected ratio of 1 : 1.* Among these 30 individuals we find:

(*a*) 11 affected and 2 unaffected males,

(*b*) 8 affected and 9 unaffected females.

There is thus a considerable preponderance of piebalds among the males.†

In conclusion I should like to express my thanks to Dr Cockayne for his helpful criticism.

REFERENCES

Cockayne, E. A. (1914). *Biometrika*, **10**, 197.

—— (1933). *Inherited Abnormalities of the Skin and its Appendages.* Oxford Univ. Press; Humphrey Milford.

—— (1935). *Biometrika*, **27**, 1.

Harman, Bishop (1909). *Trans. Ophth. Soc.* **39**, 25.

* The departure of 19 from the expected value 15 has a standard error of 2·74, and therefore cannot be considered significant.

† 11 differs from the expected value of 6·5 by 4·5; the appropriate standard error is 1·80 and the difference may therefore be significant.

A NEW METHOD OF EXPERIMENTAL SAMPLING ILLUSTRATED ON CERTAIN NON-NORMAL POPULATIONS

By G. B. HEY

1. Introduction

The theoretical distribution of many statistics calculated from small samples is known when the population is normal, but when it is not normal we know very little about the distribution of such statistics. Such work as has been done has generally assumed population forms of standard types, but we may occasionally come up against samples from populations which do not appear to fit into any known type. This has led to many attempts being made to build up, by experimental sampling from non-normal data, partial populations of samples from which can be inferred in an empirical way the laws of distribution followed by derived statistics. A list of papers dealing with this subject which have come to the author's notice is given on pp. 79, 80 below.

In many cases it has been found that in sampling from curves with one mode not at the end of the range, the distributions of statistics such as "Student's" "t", the correlation coefficient and, in certain cases, Fisher's "z", differ very slightly from one population to another. On the whole these investigations have suggested that in such cases we can neglect the departure of the population from normality without introducing serious error into our tests of significance.

The possibility of further theoretical work must not be overlooked, but unless our results are independent of population form (as, for instance, in recent work by Pitman and Welch) it is unlikely that we shall be able to make much practical use of the results. It is customary to designate a non-normal population by the values of β_1 and β_2; but in the case of samples of 100 or less from a normal population the range of values of β_1 and β_2 excluding 5 % of the total at each end is comparable with the range of β_1 and β_2 in the non-normal populations which have been used for sampling experiments. Further, this range of populations is considered by E. S. Pearson to cover most cases which will be found to occur in practice. On these grounds I think that conclusions of practical value are most likely to be reached by further sampling.

No attempt appears to have been made to carry out an experimental sampling from a bivariate population in which the distribution surface is not normal and in which the correlation coefficient is high, or to take sets of samples from a univariate non-normal population and to assign the samples to blocks and treatments in a randomized block experiment, taking a completely fresh

sample each time. Eden & Yates (1933) carried out sampling from a set of 32 values, assigning blocks in a constant arrangement and treatments at random within these blocks. Unfortunately their set of values was "as nearly normal as could be expected in a sample of 32" (Neyman, 1935, p. 114). E. S. Pearson (1931*a*) took many samples but with only one form of classification, though he suggested the consequences that were likely to follow in more complex analysis. These two papers seem to be the only ones dealing with sampling in the case of Analysis of Variance, and neither covers the state of affairs that we are considering here.

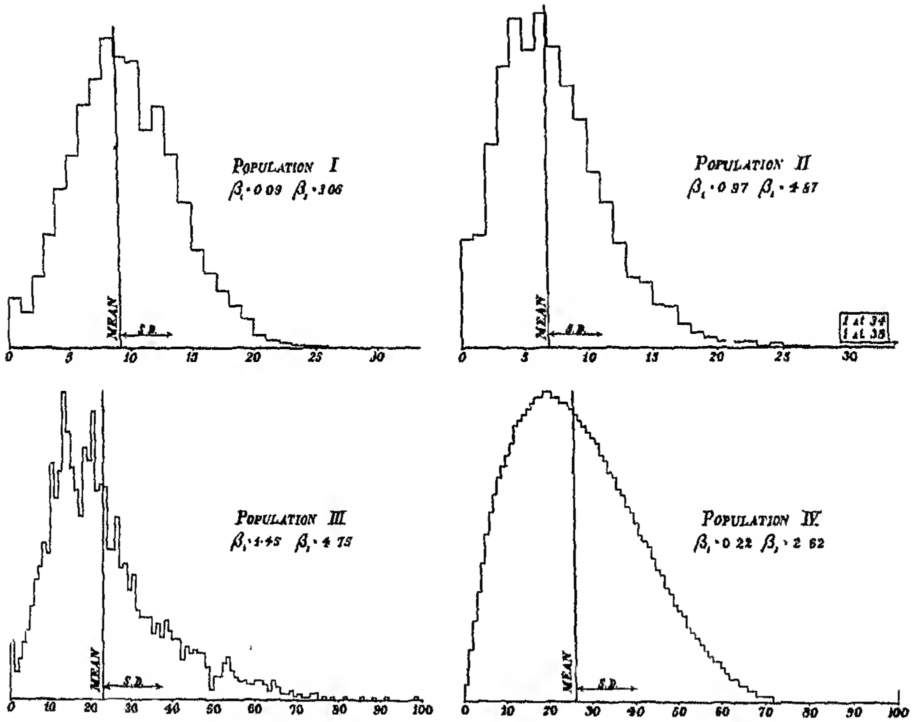
I have therefore carried out an experimental sampling from four non-normal populations of which three occurred in the course of an agricultural trial. The statistics which I have considered are the correlation and regression coefficients and the ratio of two independent estimates of variance. Now most sampling investigations have been concerned with artificial populations such as the rectangular, triangular, normal and the various Pearson types, so that the mathematical form of the frequency distribution is known. The three populations which occurred in practice did not appear to follow any mathematical law of the type usually considered, although many attempts at curve fitting were made. The populations are similar to one used by E. S. Pearson (1931*a*, *b*) but are rather more extreme and somewhat irregular.

2. Description of the experimental work

Preliminary considerations. Before commencing a practical investigation it is necessary to consider the number of samples which must be taken in order that we may have a reasonable chance of obtaining information of value. In applications to practical work we are usually interested in the tails of our derived frequency distributions—those areas at the ends of the range containing 1% or 5% of the total area under the curve. Now the chance of any sample giving a value of our statistic which lies within one of these extreme classes is small, so that the number in one of these classes is distributed approximately in the Poisson distribution.

Suppose that we agree to regard the number in one of these classes as being significantly different from expectation if it, or a more extreme value, should occur less than once in 20 times; then if our expectation is 10 we shall accept all frequencies which actually occur if they lie between 4 and 17; if the expectation is 30 the limits are about 20 and 43, and if it is 50 they are about 37 and 66. This means that if we want to be fairly sure of getting an estimate within about 25% of its value of the frequency in any class, then the expectation in that class must be about 50. We see from this that a sample of 200 (values of our statistic) will give no information about the 1% points and little about the 5% points; a sample of 1000 will give little about the 1% points and a reasonable estimate of the 5% points. Now to take 1000 random samples of size 20, say, from a population, even when using Tippett's numbers, is a considerable undertaking, and

in the case of an analysis of variance with two or more classifications the subsequent computing can be very laborious without special machines. I have therefore devised methods of doing all this automatically with the use of tabulating machines, and I believe that the method is new. The essential processes are described in a paper on the subject by Comrie, Hey and Hudson (1937). In addition to the machines there described I have used the rolling total tabulator, the main property of which is that it can transfer numbers from one counter to any other, or to any combination of others. It is very convenient for the production of the sums of squares and products of small groups of numbers. The speed of the machine is great; for instance, it produces n, Σx, Σy, Σx^2 and Σxy where there are 20 pairs (x, y) in 20 seconds.



3. Description of the populations

The populations used are shown in Table I and in the figure with the values of mean, variance, β_1 and β_2. Population I is ungrouped, being the number of ears in each of 7200 6-inch single-row lengths of wheat. No. II is grouped into intervals of 1 gram, the observations being the weights of grain on these same 7200 6-inch lengths measured to 0·1 gram; the original figures were used in the calculations, the grouping being for purposes of description only. No. III is similar to No. II,

TABLE I

Frequency distributions of the original populations

x	I	II	III			IV		
0	113	245	19	—	—	3	—	—
1	81	255	8	27	6	9	45	5
2	163	461	12	25	5	14	43	4
3	259	636	17	25	3	19	42	4
4	365	743	21	25	6	24	40	3
5	440	673	28	20	4	29	39	3
6	556	755	36	23	2	33	37	2
7	614	679	40	19	2	36	36	2
8	706	591	47	24	3	39	34	1
9	664	516	44	21	1	41	33	1
10	655	398	71	18	2	43	31	1
11	498	329	61	18	1	45	29	1
12	551	238	69	12	1	48	28	0
13	441	160	93	16	2	49	27	0
14	332	148	81	14	2	50	25	0
15	223	98	70	15	0	51	23	0
16	170	93	63	14	1	52	22	0
17	131	54	55	14	0	53	20	0
18	95	34	76	10	1	53	19	0
19	68	22	72	3	0	54	17	0
20	29	14	87	7	0	54	16	0
21	17	14	63	7	1	53	15	0
22	11	14	65	11	0	53	13	0
23	8	5	64	13	0	52	12	0
24	4	10	45	9	0	52	11	0
25	4	5	45	7	1	51	10	0
26	0	5	55	6	0	50	9	0
27	0	1	40	6	0	49	8	0
28	0	2	41	5	1	48	7	0
29	1	1 at 34	33	5	1 at 91	47	7	0
30	1	1 at 38	38	6	1 at 98	46	6	0

The second and third columns of populations III and IV give frequencies for the groups $x = 31$–60 and $x = 61$–90 respectively.

Frequency constants

Population ...	I	II	III	IV
Mean	9·172	6·838	22·949	26·157
Variance	18·07	17·78	206·0	203·2
β_1	0·089	0·97	1·446	0·219
β_2	3·058	4·87	4·755	2·625
Grouping interval	1·0	1·0	1·0	1·0

but refers to a second year's experiment, and No. IV is a smooth Pearson Type I curve whose equation is

$$y = 53{\cdot}55\left(\frac{4}{3}-\frac{x}{60}\right)^3\left(\frac{x}{20}\right).$$

The total frequency of Nos. III and IV is 2031, it being considered that a population with this total frequency was large enough.

The 7200 observations in No. III were grouped into groups of 0·3 gram, and the numbers in each group reduced in the same ratio. The correlation between the variables in populations I and II, as estimated from 7200 pairs, is 0·712, and the coefficient of regression of grain weight on ear number is 0·736, the regression being sensibly linear. The corresponding pairs for these were already punched on the same cards, and so we are taking samples from a correlation table with 7200 entries. Populations III and IV were also in pairs on the cards, but were entered so as to be uncorrelated. The methods of entering the numbers on to the cards will not be described here.

4. The calculations made on the tabulator and multiplying punch

The 7200 cards containing the first two populations were sorted into a random order and about 165 sets of 12 counted out by hand. Let us call the two numbers on each card F and G. Then the cards were passed through the multiplying punch which formed ΣF^2 and ΣFG at one run, and ΣG^2 and ΣFG at the next run; the recurrence of ΣFG provides a check. The tabulator gave the sums ΣF and ΣG (the summation is over 12 pairs). This sampling was done twice to give in all 332 samples of 12. The populations III and IV were on a new set of cards and from this set samples of 20 were taken, each 20 being replaced before the next 20 were drawn, until 1008 samples had been taken from each population. A complete list of these is given by the tabulator, together with the totals of each set of 20.

Using the rolling total tabulator we produce twice the sum of squares of the numbers in each sample, and at the same time twice the sum of products of the two numbers, one from each population, and also the sum of the 20 numbers themselves. By this means we have all the totals that we require and several checks on the operations of the machine. The sums of squares are all hand-punched on to new cards. We must next assign the four imaginary blocks and five imaginary treatments to the 20 numbers; this is done by identifying the cards with 11, 12, 13, 14, 15; 21, 22, 23, 24, 25; 31, ..., 35; 41, ..., 45. It is now possible with a single run through the tabulator to produce the sums of the numbers in groups according to the first or second figure of the identification. These sums are referred to as the block and treatment totals, or the sums in fours or fives. The totals in fours, fives and twenties are hand-punched on to further cards and sorted into groups so that all equal numbers are together, and

it is arranged that each group is preceded by a card containing the square of that number. This square is transferred by the reproducing punch to each card of the following group until the group ends, when the number which is being transferred is changed automatically. We have now certain figures available for constructing the Analysis of Variance table. Calling the numbers x_{ij}, $(i=1,2,3,4;\ j=1,2,3,4,5)$, we have:

$$\sum_j x_{ij}=B_i \qquad \sum_i\sum_j x_{ij}=G$$
$$\sum_i x_{ij}=T_j \qquad \sum_i\sum_j x_{ij}^2=S$$

and our Analysis of Variance will read:

Variation due to	D.F.	Sum of squares	
Blocks	3	$\frac{1}{5}(\Sigma B_i^2)-\frac{1}{20}G^2$	$=\frac{1}{20}[4\Sigma B_i^2-G^2]$
Treatments	4	$\frac{1}{4}(\Sigma T_j^2)-\frac{1}{20}G^2$	$=\frac{1}{20}[5\Sigma T_j^2-G^2]$
Error	12	$S-\frac{1}{5}(\Sigma B_i^2)-\frac{1}{4}(\Sigma T_j^2)+\frac{1}{20}G^2$	$=\frac{1}{20}[20S-4\Sigma B_i^2-5\Sigma T_j^2+G^2]$
TOTAL	19	$S-\frac{1}{20}G^2$	$=\frac{1}{20}[20\,S-G^2]$

Now we cannot make the tabulator divide; we can make it multiply by two, and by repeating this process and combining the contents of counters in various ways it is possible to produce all the quantities in square brackets in the table, and since we shall be concerned only with the ratios of these quantities, we can neglect the factor 20. By feeding the cards containing B_i^2, T_j^2, G^2 and S, and by a suitable arrangement of counters, the machine produces the table in the form shown above in less than five seconds.

5. SUBSEQUENT CALCULATIONS

Since the tabulator cannot divide other than, in effect, by multiplying by the least common multiple, it is impossible for us to get any further using the automatic machines. However, the work which has been done on them has resulted in an immense saving of labour on a very dull task.

First experiment. We set ΣF^2 on the levers of a Brunsviga and multiply it by 12, and then after subtracting $(\Sigma F)^2$, we have $12\Sigma\,(F-\bar{F})^2$; carrying out similar calculations for ΣFG and ΣG^2 we produce the correlation and regression coefficients very rapidly. The distribution of the correlation coefficient and of Fisher's transformation of this coefficient are considered; also the distribution of the regression coefficient is worked out. In addition to this the sets of samples of 12 for population I were combined by taking 6 at random, and from this set of 72 we can form estimates of the variance within and between the sets of 12. The ratio of these estimates, based respectively on 5 and 66 degrees of freedom, is distributed in a known form derived from the Incomplete Beta Function when the population is normal. Altogether 383 such samples were taken. The results, which diverged considerably from expectations and led up to the second experiment, are examined in the next section.

Second experiment. We have three estimates of the variance in the original population based on 3, 4 and 12 degrees of freedom, and we know, in the case of the original population being normal, the distribution of the ratios of these estimates. (Notice that although the third ratio is not independent of the other two, the three distributions are independent in the case of normal data.) The distributions of these three ratios were drawn up; the manner of doing this was to find the 1, 5, 10, 20, 40, 60, 80, 90, 95 and 99 % points in the case of samples from normal data and to count the numbers of samples giving values lying within these classes. The theoretical values are calculated from the tables of the Incomplete Beta Function by inverse interpolation; a check is given for the two end classes from the tables of Fisher's "z". This interpolation was very difficult in places owing to the large tabular interval, and in certain cases the tables had to be recomputed at a smaller interval.

Finally the correlation between the 20 pairs of values from the two populations was evaluated for 336 sets of 20, and the distribution of totals of five for each population, and its first four moments, produced. We are now in a position to discuss the results of these calculations.

6. Discussion of the results

First experiment. The observed distribution of the transformed correlation coefficient, $z=\frac{1}{2}\log_e(1+r)/(1-r)$, shown in Table II, has mean 0·988 after allowing for the bias, S.D. = 0·349, $\beta_1=0{\cdot}057$, and $\beta_2=3{\cdot}275$, the expected values on the basis of normal theory being 0·892, 0·328, 0, 3·22, using the expressions given by Fisher (1923). β_1 is almost significantly, and the variance and β_2 insignificantly, different from expectation. The difference in means is 0·096 and is very significant, its standard error being 0·018. On the whole the distribution agrees fairly closely with expectation except for this shift in the mean, the cause of which remains doubtful. There is no doubt of the correctness of the value 0·712 for the correlation as calculated from 7200 pairs, but it is interesting to note that, if we combine the observations in pairs according to their position in the field, the correlation between the 3600 pairs of totals of grain weight and ear number becomes 0·667. In the work done on the experiment in which these figures occurred the totals were combined in 32 different ways, and the correlation was estimated from the totals of larger units and also from the figures for 6-inch lengths within the larger units, and in the latter case the correlation was steady at about 0·76, as estimated from the "within plots" line of the Analysis of Covariance, if the number of 6-inch lengths in the plot was more than four. If we take this value as being the correlation between the two counts, then our sampling experiment agrees closely with expectation, as based on normal theory.

The distribution of the regression of G on F has one observation far removed from the rest, and this has been omitted in forming estimates of the parameters of the distribution. (It is about 6σ from the mean and this is very unlikely in a

sample of 332.) The mean is 0·735, s.d. = 0·245, $\beta_1 = 0{\cdot}001$, $\beta_2 = 3{\cdot}51$, with expectations 0·736, 0·234, 0, 3·86 (K. Pearson, 1926, p. 7). These values agree well with expectation. The purpose of the sampling being primarily to test the effect of non-normality on the distribution of the correlation coefficient, it was felt that 332 samples would give sufficiently accurate information, since we do

TABLE II

Sampling distributions in Experiment 1

Transformed correlation coefficient		Correlation coefficient		Regression coefficient of G on F	
z	Frequency	r	Frequency	Coefficient	Frequency
−0·4 to −0·3	1	At −0·46	1	−0 2 to −0·1	1
−0·3–	0			−0·1–0·0	0
−0·2–	0	At −0·07	1	0·0–0·1	2
−0·1–	1			0·1–	4
0·0–	0	0·10–0·14	1	0·2–	3
0·1–	4	0·14–	2	0·3–	16
0·2–	1	0·18–	1	0·4–	22
0·3–	3	0·22–	1	0·5–	48
0·4–	11	0·26–	1	0·6–	57
0·5–	14	0·30–	2	0·7–	47
0·6–	18	0·34–	0	0·8–	45
0·7–	39	0·38–	6	0 9–	44
0·8–	34	0·42–	5	1·0–	22
0·9–	31	0·46–	5	1·1–	11
1·0–	35	0·50–	10	1·2–	3
1·1–	37	0·54–	7	1·3–	5
1·2–	30	0·58–	17	1·4–	0
1·3–	25	0·62–	29	1·5–1 6	1
1·4–	19	0·66–	27	...	...
1·5–	11	0·70–	24	...	...
1·6–	13	0·74–	30	2·2–2·3	1
1·7–	4	0·78–	40		
1·8–1·9	1	0·82–	48		
		0·86–	40		
		0·90–	29		
		0·94–0·98	5		
Total	332	Total	332	Total	332

not usually require great accuracy in a measure of association. The observed distribution of the transformed r fits a normal curve quite well, and there is no evidence of lack of agreement at the tails. All this suggests that considerable non-normality in the original distribution will not affect the distributions of correlation and regression coefficients even in the case of high correlation.

The results of the comparison of estimates of variance is to give frequencies 4, 12, 16, 31, 59, 63, 81, 59, 23, 26, 9, in classes with expectations 3·83, 15·3,

19·1, 38·3, 76·6, 76·6, 76·6, 38·3, 19·1, 15·3, 3·83, the total frequency being 383. There is a serious excess of observed values at the end of this distribution, but since we are using the same sets of 12 several times this may cause a bias, and the more extensive second experiment was carried out to examine this case more closely.

Second experiment. The first step is to test the randomness of the sampling process, since this is new. Owing to the method used the results come out in nine batches of 112 and for each batch separately we have the mean and variance of the distribution of totals of fives. These were found to agree well with expectation with two exceptions, one of which was due to an oversight in sampling for one batch of population IV; this batch was discarded. One batch from population III was insufficiently variable, but there was no obvious reason for this, and the batch was retained. The first four moments of the distribution of totals of five combined from the eight (or nine) batches, together with their expected values, taking into account the non-normality, are shown in Table III.

TABLE III

	Mean	Variance	β_1	β_2
Population III:				
Expected	114·7	1030	0·289	3·351
Observed	114·2	960	0·323	3·347
Population IV:				
Expected	130·6	1015	0·044	2·925
Observed	130·7	1027	0·072	2·915

The variance of population III is significantly smaller than expectation, its S.D. being approximately 22. The other values agree closely with expectation, using the approximate values of the S.D. of β_1 and β_2. The S.D. of each of the means is about 0·5.

Further, the correlation between the two values from populations III and IV was evaluated for 336 pairs and found to be $-0{\cdot}0127$ with S.D. of approximately 0·016. The manner in which the two populations were paired is the same as that used for the actual sampling, and was intended to produce zero correlation between the two variables; we can therefore consider that the adequacy of this method of sampling has been demonstrated.

The observed frequencies of the ratios of variances* in the classes 0–1 %, 1–5 %, etc. are given in Table IV, together with the expectations in those classes based on normal theory. The agreement in all cases is good, and with one exception there is no evidence of serious divergence at the tails. The shortage in the 1 % class for population III, degrees of freedom 3 : 12, is significant by itself,

* Taken from the Analysis of Variance table given on p. 73 above.

TABLE IV

Distribution of ratios of estimates of variances in Experiment 2

Population III

Class %	Observed frequencies: Degrees of freedom 3 : 4	3 : 12	4 : 12	Normal theory expectation
0– 1	6	1	8	10·08
1– 5	36	38	40	40·32
5– 10	49	46	59	50 40
10– 20	78	93	90	100·80
20– 40	206	209	204	201·60
40– 60	211	203	212	201·60
60– 80	189	199	199	201·60
80– 90	114	98	99	100·80
90– 95	66	56	49	50·40
95– 99	41	52	41	40·32
99–100	12	13	7	10·08
Totals	1008	1008	1008	1008·00
χ^2	11·5	4·57	0·86	

Population IV

Class %	Observed frequencies: Degrees of freedom 3 : 4	3 : 12	4 : 12	Normal theory expectation
0– 1	12	9	7	8·96
1– 5	43	43	42	35·84
5– 10	48	41	41	44·80
10– 20	97	79	81	89·60
20– 40	179	169	160	179·20
40– 60	160	180	164	179·20
60– 80	173	190	180	179·20
80– 90	88	74	118	89·60
90– 95	48	44	52	44·80
95– 99	35	58	38	35·84
99–100	13	9	13	8·96
Totals	896	896	896	896·00
χ^2	4·81	1·71	13·5	

but it is the lowest of six values and there does not appear to be any trend over the rest of the range. The six values of χ^2, each based on a classification into 5 groups, were evaluated and are 11·5, 4·57, 0·86, 4·81, 1·71 and 13·5. This is quite a reasonable set for 4 degrees of freedom.

7. Conclusion

Samples have been taken from four non-normal populations and the distributions of correlation coefficients, regression coefficients, and the ratio of different estimates of variance corresponding to degrees of freedom 3:4, 3:12 and 4:12 have been found. They all agree sufficiently well with the known distributions in the case of normal populations for us to neglect the departure from normality in using these tests of significance when the original populations are of the form we have used. This agrees with the general conclusions reached by E. S. Pearson in other cases of sampling from non-normal populations. The bias found in the first set of ratios of estimated variances may be due to the dependence among samples as a result of using the same sets of 12 over and over again.

Now that we have a method of taking random samples and of doing most of the subsequent computing automatically it would be of considerable interest to continue the sampling investigations for the further investigation of the ratios of estimates of variance in the cases of multiple classification, e.g. the Latin Square and multiple factor experiments, and for other forms of population. All these can be rapidly carried out with the aid of tabulating machines.

Finally I wish to thank Dr J. Wishart for his valuable advice and continued interest taken in this work; also Mr J. Mandeville of the British Tabulating Machine Co., Ltd., and Dr L. J. Comrie for their assistance in connection with the parts of the work involving tabulating machines.

REFERENCES

Comrie, L. J., Hey, G. B. & Hudson, H. G. (1937). *J.R. Statist. Soc.* Supp. **4**, 210.
Eden, T. & Yates, F. (1933). *J. Agric. Sci.* **23**, 6.
Fisher, R. A. (1923). *Metron*, **1**, part 4, 3.
Neyman, J. (1935). *J. R. Statist. Soc.* **2**, 108.
Pearson, E. S. (1931*a*). *Biometrika*, **23**, 114.
—— (1931*b*). *J. Amer. Statist. Ass.* **26**, 128.
Pearson, Karl (1926). *Proc. Roy. Soc.* A, **112**, 7.

BIBLIOGRAPHY

The following list, while not claiming to be exhaustive, contains the chief papers with which the author is familiar dealing with artificial sampling experiments (particularly those relating to non-normal populations) and certain relative papers.

1908. "Student." "The probable error of a mean." *Biometrika*, **6**, 1.

"Student." "The probable error of a correlation coefficient." *Biometrika*, **6**, 302.

1925. Tippett, L. H. C. "On the extreme individuals and the range of samples taken from a normal population." *Biometrika*, **17**, 364.

Neyman, J. "Contributions to the theory of small samples drawn from a finite population." *Biometrika*, **17**, 472.

1926. Church, A. E. R. "On the mean squared standard deviations of small samples from any population." *Biometrika*, **18**, 321.

1928. Shewhart, W. A. & Winters, F. W. "Small samples—new experimental results." *J. Amer. Statist. Ass.* **23**, 144.

Neyman, J. & Pearson, E. S. "On the use and interpretation of certain test criteria for purposes of statistical inference." *Biometrika*, **20A**, 175, 263.

Pearson, E. S. & Adyanthāya, N. K. "The distribution of frequency constants of small samples from symmetrical populations." *Biometrika*, **20A**, 356.

Holzinger, K. & Church, A. E. R. "On the means of samples from a U-shaped population." *Biometrika*, **20A**, 361.

"Sophister." "Discussion of small samples drawn from an infinite skew population." *Biometrika*, **20A**, 389.

1929. Baker, G. A. "Random sampling from non-homogeneous populations." *Metron*, **8**, 67.

Rider, P. R. "On the distribution of the ratio of mean to standard deviation in small samples from non-normal universes." *Biometrika*, **21**, 124.

Pepper, J. "Studies in the theory of sampling." *Biometrika*, **21**, 231.

Pearson, E. S. & Others. "The distribution of frequency constants in small samples from non-normal symmetrical and skew populations." *Biometrika*, **21**, 259.

Pearson, E. S. "Some notes on sampling tests with two variables." *Biometrika*, **21**, 337.

1931. Pearson, E. S. "The test of significance for the correlation coefficient." *J. Amer. Statist. Ass.* **26**, 128, 424.

Rider, P. R. "On small samples from certain non-normal universes." *Ann. Math. Statist.* **2**, 48.

Dunlap, H. F. "An empirical determination of the distribution of means, standard deviations, and correlation coefficients drawn from rectangular populations." *Ann. Math. Statist.* **2**, 66.

Baker, G. A. "The relation between the mean and variance, mean squared and variance in samples from combinations of normal populations." *Ann. Math. Statist.* **2**, 333.

Pearson, E. S. "The analysis of variance in cases of non-normal variation." *Biometrika*, **23**, 114.

Roux, J. M. Le. "A study of the distribution of variance in small samples." *Biometrika*, **23**, 134.

1932. Carlson, J. L. "A study of the distribution of means estimated from small samples by the method of maximum likelihood for Pearson's type II curves." *Ann. Math. Statist.* **3**, 86.

Georgescu, N. St. "Further contributions to the sampling problem." *Biometrika*, **24**, 65.

Pearson, K. "Experimental discussion of the ($\chi^2.P$) test for goodness of fit." *Biometrika*, **24**, 351.

1932. RIDER, P. R. "On the distribution of the correlation coefficient in small samples." *Biometrika*, **24**, 382.

EDEN, T. & YATES, F. "On the validity of Fisher's z-test when applied to an actual example of non-normal data." *J. Agric. Sci.* **23**, 6.

CHESIRE, L., OLDIS, E. & PEARSON, E. S. "Further experiments on the sampling distribution of the correlation coefficient." *J. Amer. Statist. Ass.* **27**, 121.

1933. CRAIG, A. T. "On the correlation between certain averages from small samples." *Ann. Math. Statist.* **4**, 127.

ROBINSON, S. "An experiment regarding the χ^2 test." *Ann. Math. Statist.* **4**, 285.

PERLO, V. "On the distribution of Student's ratio for samples of three drawn from a rectangular distribution." *Biometrika*, **25**, 203.

1934. BAKER, G. A. "Transformation of non-normal frequency distributions into normal distributions." *Ann. Math. Statist.* **5**, 113.

HANSMANN, G. H. "On certain non-normal symmetric frequency distributions." *Biometrika*, **26**, 129.

1937. PEARSON, E. S. & WELCH, B. L. "Notes on some statistical problems raised in Mr Bayes paper (on variability of cotton cloth strength)." *J.R. Statist. Soc.* Supp. **4**, 94.

PITMAN, E. J. G. "Significance tests which may be applied to samples from any population." *J.R. Statist. Soc.* Supp. **4**, 119, 225.

WELCH, B. L. "On the z-test in randomized blocks and Latin squares." *Biometrika*, **29**, 21.

PEARSON, E. S. "Some aspects of the problem of randomization." *Biometrika*, **29**, 53.

A NEW MEASURE OF RANK CORRELATION

By M. G. KENDALL

1. In psychological work the problem of comparing two different rankings of the same set of individuals may be divided into two types. In the first type the individuals have a given order A which is objectively defined with reference to some quality, and a characteristic question is: if an observer ranks the individuals in an order B, does a comparison of B with A suggest that he possesses a reliable judgment of the quality, or, alternatively, is it probable that B could have arisen by chance? In the second type no objective order is given. Two observers consider the individuals and rank them in orders A and B. The question now is, are these orders sufficiently alike to indicate similarity of taste in the observers, or, on the other hand, are A and B incompatible within assigned limits of probability? An example of the first type occurs in the familiar experiments wherein an observer has to arrange a known set of weights in ascending order of weight; the second type would arise if two observers had to rank a set of musical compositions in order of preference.

The measure of rank correlation proposed in this paper is capable of being applied to both problems, which are, in fact, formally very much the same. For purposes of simplicity in the exposition it has, however, been thought convenient to preserve a distinction between them.

Definition of τ

2. Consider a set of individuals, numbered from 1 to 10, whose objective order is that of the natural sequence 1, 2, 3, ..., 10, and consider an arbitrary ranking such as the following:

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

Consider the order of the nine pairs of numbers obtained by taking the first number 4, with each succeeding number. The first pair, 4 7, is in the correct order (in the sequence of 1, 2, ..., 10), and we therefore allot it a score $+1$. The second pair, 4 2, is in the wrong order and we score -1. The third pair, 4 10, scores $+1$, and so on, the nine scores being

$$+1-1+1-1+1+1-1+1+1, \text{ totalling } +3.$$

Consider also the scores of the second number, 7, with its eight succeeding numbers. They are

$$-1+1-1-1+1-1-1+1, \text{ totalling } -2.$$

The scores of the third number are

$$+1+1+1+1-1+1+1, \text{ totalling } +5.$$

Proceeding thus with each number, we have 9 scores, as follows

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1.$$

The total of these scores is $+5$.

Now the maximum score, obtained if the numbers are all in the objective order (1, 2, ..., 10), is 45. I therefore define a rank correlation coefficient between a variable ranked in the objective order (1, 2, ..., 10) and the variable ranked in the order above as

$$\tau = \frac{\text{actual score}}{\text{maximum possible score}} = \frac{5}{45} = +0{\cdot}11.$$

Generally, if there are n individuals, the maximum score, obtained if and only if they are all in objective order is $(n-1)+(n-2)+\ldots+1 = \frac{n(n-1)}{2}$. Denoting the actual score for any given ranking by Σ, we may calculate a measure of the rank correlation between this ranking and the objective ranking by putting

$$\tau = \frac{2\Sigma}{n(n-1)}. \qquad \ldots\ldots(1)$$

TWO SHORT METHODS FOR THE CALCULATION OF τ

3. τ is calculable more easily than might appear at first sight from the above approach. Consider for example the order given above, viz.

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

We see that the number 1 has two numbers on its right and 7 on its left. We therefore score $+2-7=-5$, and then strike out the 1, being left with

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 5 \quad 9$$

The number 2 has 6 numbers on its right and two on its left and hence we score $6-2=+4$. We then strike out the 2 and proceed with the 3 and so on. It will be found that the scores obtained are

$$-5, \quad +4, \quad +1, \quad +6, \quad -3, \quad 0, \quad +3, \quad 0, \quad -1.$$

The total of these scores is $+5$, and is equal to Σ.

The above rule is quite general. Its validity will be evident when it is noted that instead of taking the first number with each succeeding number and so on, as in § 2, we consider the pairs contributing to Σ in a different way. Taking the number 1 first, and remembering that all the other numbers are greater than 1, we see that any number on the left must contribute -1 to Σ and any number on the right contributes $+1$. When 1 is struck out the procedure remains valid for 2, and so on.

4. Alternatively, the following procedure may be adopted: Considering once again the order

$$4 \quad 7 \quad 2 \quad 10 \quad 3 \quad 6 \quad 8 \quad 1 \quad 5 \quad 9$$

we see that the first number, 4, has on its right 6 numbers which are greater. The second number, 7, has on its right 3 numbers which are greater. The third number, 2, has on its right 6 numbers which are greater; and so on. The numbers so obtained are

$$6, \quad 3, \quad 6, \quad 0, \quad 4, \quad 2, \quad 1, \quad 2, \quad 1$$

totalling 25.

There must, therefore, be $45-25 = 20$ numbers lying to the right of successive numbers in the order which are less than those numbers, and hence

$$\begin{aligned} \Sigma &= 25-20 \\ &= +5, \text{ as before.} \end{aligned}$$

Generally, if the number obtained by the above method of counting greater numbers is k

$$\Sigma = 2k - \frac{n(n-1)}{2}.$$

In practice, I find this method convenient and rapid. It has, moreover, the advantage of providing an independent check; for if the process is repeated counting greater numbers which lie *to the left*, giving a total of, say, l,

$$\Sigma = \frac{n(n-1)}{2} - 2l.$$

5. The use of τ can now be extended to the case where no objective order exists. In fact, given two rankings, A and B, of the same set of individuals τ may be defined as the coefficient obtained by regarding one order, A, as an objective order. If, for example, the orders are as follows:

A	6	9	4	3	5	10	2	1	8	7
B	6	5	10	2	3	9	7	4	1	8

τ is given by first rearranging A as an objective order, writing below it the corresponding member in B, thus

A'	1	2	3	4	5	6	7	8	9	10
B'	4	7	2	10	3	6	8	1	5	9

and then calculating τ in the manner of preceding paragraphs. Actually, as will be seen below, it is not necessary in any practical calculations to rewrite the orders in this way.

6. It is a notable fact that the same coefficient τ is reached whichever of the two orders, A and B, is rearranged as an objective order.

Consider again the orders given in the preceding paragraph, namely,

A'	1	2	3	4	5	6	7	8	9	10
B'	4	7	2	10	3	6	8	1	5	9

Rearranging B as an objective order we have

A''	8	3	5	1	9	6	2	7	10	4
B''	1	2	3	4	5	6	7	8	9	10

If we repeat this operation on the A'' and B'' we shall get back to A' and B'. A', B' and A'', B'' are thus reciprocally related and the permutations B' and A'' may be said to be *conjugate*.

We have to show that τ is the same when calculated from B' when A' is the objective ranking as when calculated from A'' when B'' is the objective ranking, i.e. that Σ is the same for two conjugate permutations with regard to an objective order 1, 2, ..., n.

In § 2, the value of Σ for B' was ascertained directly, the various items entering into the sum being

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1.$$

Consider now the value of Σ for A'' *obtained by the short method of* § 3.

The sums entering into Σ will be found to be

$$+3, \quad -2, \quad +5, \quad -6, \quad +3, \quad 0, \quad -1, \quad +2, \quad +1,$$

i.e. exactly the same as those for B' obtained by the more direct method; and hence Σ and τ are the same in the two cases.

This result is true in general. If the permutation B' begins with a number a_0 the contribution to $\Sigma_{B'}$ from pairs involving a_0 will be $(n-a_0)-(a_0-1)$. In A'' the a_0th number will be 1 and the contribution to $\Sigma_{A''}$ will also be $(n-a_0)-(a_0-1)$, in the manner of § 3. If the second number in B' is a_1 the contribution to $\Sigma_{B'}$ will be $(n-a_1)-(a_1-1)\pm 1$ according to whether a_1 is greater than a_0 or not. In A'' the a_1th number will be 2, and the contribution to $\Sigma_{A''}$ is also $(n-a_1)-(a_1-1)\pm 1$ according to whether 1 lies to the left or the right of 2 in A'', i.e. whether a_1 is greater than a_0 or not; and so on.

Thus Σ and τ are the same for two conjugate permutations with regard to the objective order 1, 2, ..., n.

7. In practical cases, the value of τ may be found as follows:

Write down above the given rankings the objective ranking. In the example already considered this would give

	1	2	3	4	5	6	7	8	9	10
A	6	9	4	3	5	10	2	1	8	7
B	6	5	10	2	3	9	7	4	1	8

The number 1 in B has an 8 above it in A. In the objective ranking 8 has two numbers to the right and seven to the left. Score, therefore, -5 and strike out the 8 in the objective ranking. The number 2 in B has a 3 above it in A, and 3 in the objective ranking has six numbers to its right (ignoring the number struck out) and two to its left, score $+4$; and so on, the scores being

$$-5, \quad +4, \quad +1, \quad +6, \quad -3, \quad 0, \quad +3, \quad 0, \quad -1,$$

totalling $+5$, which is equal to Σ.

8. τ satisfies certain elementary requirements of a measure of rank correlation. It is $+1$ if and only if correspondence between the rankings of A and B is perfect. It is -1 if and only if the rankings are exactly inverted. For intermediate values it appears to provide a satisfactory measure of the correspondence between the two rankings. A few examples for $n = 10$ will give some idea of the scale of measurement which it provides (an objective order 1, 2, ..., 10 is taken in each case):

Order	τ	ρ*
4 7 2 10 3 6 8 1 5 9	+0·11	+0·14
1 6 2 7 3 8 4 9 5 10	+0·56	+0·64
7 10 4 1 6 8 9 5 2 3	−0·24	−0·37
6 5 4 7 3 8 2 9 10 1	+0·02	+0·03
10 1 2 3 4 5 6 7 8 9	+0·60	+0·45
10 9 8 7 6 1 2 3 4 5	−0·56	−0·76

In the case where no objective ranking exists τ measures the closeness of correspondence between two given rankings in the sense that it measures how accurate either ranking would be if the other were objective. In other words it measures the *compatibility* of two rankings.

9. For the purpose of measuring correlation between ranks, therefore, τ appears to compare favourably with ρ. It is admitted that ρ can take $\frac{n^3-n}{6}$ values between -1 and $+1$, whereas τ can take only $\frac{n^2-n}{2}$ values in the range. This does not, however, appear to constitute a serious disadvantage to the sensitivity of τ.

On the other hand, τ possesses one marked advantage over ρ, in that it is not difficult to find the distribution of values obtained by correlating a given ranking with the members of a universe in which all possible rankings occur equally

* Throughout this paper ρ means the Spearman coefficient of rank correlation defined by

$$\rho = 1 - \frac{6S(d^2)}{n^3-n},$$

where d is a difference in ranks.

frequently. It is shown below that the distribution of τ tends to normality for large n, resembling ρ in this respect; but in fact τ is surprisingly close to normality even for low values of n, whereas the distribution for ρ has not yet been given, and appears to present peculiar features.*

THE SAMPLING DISTRIBUTION OF Σ

10. To judge the significance of an observed value of τ or of Σ in the case where an objective order is given, we wish to know whether the value could have arisen by chance from a universe in which all the possible rankings of the n objects occur an equal number of times. It is, therefore, necessary to consider the distribution of Σ in such a universe. The distribution of τ may be found at once from that of Σ by dividing the variate values of Σ by $\frac{n(n-1)}{2}$.

The same distribution may be used to judge the significance of a value of τ expressing the compatibility of two rankings. A significantly negative τ, for example, would mean that if one ranking is taken to be objective the other has not, as judged by the τ-distribution, arisen by chance from the universe in which all possible rankings occur equally frequently; in other words that the two rankings are significantly incompatible.

Consider then the universe of values of Σ obtained from an objective order 1, 2, 3, ..., n and the $n!$ possible permutations of the first n integers. Let the number of values of a given Σ be denoted by $u_{n,\Sigma}$. Consider a given ranking of the numbers 1, ..., n, and the effect of inserting an additional number $(n+1)$ in the various possible places in this ranking, from the first place (preceding the first member of the rank of n) to the last place (following the last member of the rank of n).

Inserting the number $(n+1)$ at the beginning will add $-n$ to the value of Σ. Inserting it between the first and second members will add $-(n-2)$ to Σ. Inserting it between the second and third will add $-(n-4)$ to Σ, and so on. Adding the number $(n+1)$ at the end will add $+n$ to Σ.

It follows that

$$u_{n+1,\Sigma} = u_{n,\Sigma-n} + u_{n,\Sigma-n+2} + u_{n,\Sigma-n+4} + \ldots + u_{n,\Sigma+n-4} + u_{n,\Sigma+n-2} + u_{n,\Sigma+n}. \qquad \ldots\ldots(2)$$

This recursion formula permits of the calculation of the frequency array of Σ.

11. If $n = 2$, there are two values of Σ, $+1$ and -1, i.e. $u_{2,-1} = u_{2,1} = 1$, $u_{2,0} = 0$. From (2) we have

$$u_{3,\Sigma} = u_{2,\Sigma-4} + u_{2,\Sigma-2} + u_{2,\Sigma} + u_{2,\Sigma+2} + u_{2,\Sigma+4},$$

* The fact that ρ tends to normality for large n has recently been proved by Hotelling & Pabst (1936). The remarks above on the behaviour of ρ for low values of n are founded on an expression for the sampling distribution of ρ which will be discussed in a further communication shortly to be published. This communication will also deal with the relation between τ and ρ.

the possible values of Σ ranging from -3 to $+3$. By substituting $\Sigma = -3, \ldots +3$ in the above, we find

$$u_{3,3} = 1, \quad u_{3,2} = 0, \quad u_{3,1} = 2, \quad u_{3,0} = 0,$$

and similar values for the negative values of Σ.

Applying equation (2) again we find

$$u_{4,6} = 1, \quad u_{4,5} = 0, \quad u_{4,4} = 3, \quad u_{4,3} = 0,$$
$$u_{4,2} = 5, \quad u_{4,1} = 0, \quad u_{4,0} = 6, \quad \text{etc.}$$

The successive arrays of Σ may in fact be built up by the following process:

$$\begin{array}{ccccccc}
1 & 1 & & & & & \\
 & 1 & 1 & & & & \\
 & & 1 & 1 & & & \\
\cline{1-4}
1 & 2 & 2 & 1 & & & \\
 & 1 & 2 & 2 & 1 & & \\
 & & 1 & 2 & 2 & 1 & \\
 & & & 1 & 2 & 2 & 1 \\
\hline
1 & 3 & 5 & 6 & 5 & 3 & 1 \\
\text{etc.} & & & & & &
\end{array}$$

At each stage, to find the array for $(n+1)$ we write down the n-array $(n+1)$ times, one under the other and moving one place to the right each time, and then sum the $(n+1)$ arrays. If the total array has a central value, that value is the frequency for $\Sigma = 0$, and all values of Σ must be even. If the total array has two central values, these values are the frequencies for $\Sigma = \pm 1$, and all values of Σ must be odd.

12. The above procedure may be condensed by forming a kind of figurate triangle as follows:

1	1										
2	1	1									
3	1	2	2	1							
4	1	3	5	6	5	3	1				
5	1	4	9	15	20	22	20	15	9	4	1
etc.	etc.										

In this array, a number in the rth row is the sum of the number immediately above it and the $(r-1)$ numbers to the left of that number. The formation of the array is quite simple and several devices shorten the arithmetic. For instance, in part of the array towards the left a number in the rth row is the sum of the number immediately above it and the number immediately to the left. A check is provided by the fact that the total in the rth row is $r!$.

The following table shows the frequency distribution of Σ for values of n from 1 to 10.

TABLE I

Distribution of Σ for values of n from 1 to 10 (only the positive half of the symmetrical distribution shown)

Values of Σ

Values of n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1														
2	0	1													
3	0	2	0	1											
4	6	0	5	0	3	0	1								
5	22	0	20	0	15	0	9	0	4	0	1				
6	0	101	0	90	0	71	0	49	0	29	0	14	0	5	0
7	0	573	0	531	0	455	0	359	0	259	0	169	0	98	0
8	3836	0	3736	0	3450	0	3017	0	2493	0	1940	0	1415	0	961
9	29228	0	28675	0	27073	0	24584	0	21450	0	17957	0	14395	0	11021
10	0	250749	0	243694	0	230131	0	211089	0	187959	0	162337	0	135853	0

Values of Σ

Values of n	15	16	17	18	19	20	21	22	23	24	25	26	27	28
6	1													
7	49	0	20	0	6	0	1							
8	0	602	0	343	0	174	0	76	0	27	0	7	0	1
9	0	8031	0	5545	0	3606	0	2191	0	1230	0	628	0	285
10	110010	0	86054	0	64889	0	47043	0	32683	0	21670	0	13640	0

Values of Σ

Values of n	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
9	0	111	0	35	0	8	0	1									
10	8095	0	4489	0	2298	0	1068	0	440	0	155	0	44	0	9	0	1

The frequency polygon of the distribution is quite close to normality even for $n = 6$. For $n = 10$ the correspondence is very good over the material part of the range, as may be judged roughly by drawing the frequency polygon to Σ and the normal curve with the same area and standard deviation. On an ordinary scale the two curves are hardly distinguishable by the eye above $\Sigma = 5$.

STANDARD ERROR OF τ

13. A little consideration of the above method of obtaining the frequency distribution of Σ will show that the distribution may be arrayed by the function:

$$f = (x^{-1}+x)(x^{-2}+1+x^2)(x^{-3}+x^{-1}+x+x^3)\ldots$$
$$(x^{-(n-1)}+x^{-(n-3)}+\ldots+x^{(n-3)}+x^{(n-1)}). \quad \ldots\ldots(3)$$

The coefficient of x^Σ in f is the frequency of Σ in the distribution.

If we differentiate f with respect to x and then multiply by x the coefficient of x^Σ is multiplied by Σ. Writing then θ for the operator $x\dfrac{\partial}{\partial x}$ we have

$$\mu_0\mu_1 = (\theta f)_{x=1},$$

and generally
$$\mu_0\mu_r = (\theta^r f)_{x=1}. \quad \ldots\ldots(4)$$

Applying equation (4) when $r = 2$, I find

$$\mu_2 = \frac{n(n-1)(2n+5)}{18}, \quad \ldots\ldots(5)$$

and hence the standard error of τ is given by

$$\sigma_\tau = \frac{1}{3}\sqrt{\frac{2(2n+5)}{n(n-1)}}, \quad \ldots\ldots(6)$$

which, as n becomes large, gives

$$\sigma_\tau \sim \frac{2}{3}\cdot\frac{1}{\sqrt{n}}. \quad \ldots\ldots(7)$$

Table II shows the proportion of the total frequencies falling outside ranges $\pm\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ for some of the distributions of Table I.

The expected values on the hypothesis of a normal distribution are 0·3173, 0·0455, 0·0027 and it is clear that for most practical purposes in testing the significance of an observed τ for $n = 10$ or greater, the standard error may be used in the ordinary way.

14. Applying equation (4) when $r = 4$, I find

$$\mu_4 = \frac{n(n-1)}{2}\left\{1+\frac{74}{9}(n-2)+\frac{37}{6}(n-2)(n-3)+\frac{32}{25}(n-2)(n-3)(n-4)\right.$$
$$\left.+\frac{2}{27}(n-2)(n-3)(n-4)(n-5)\right\}. \quad \ldots\ldots(8)$$

TABLE II

Proportion of frequencies of the distribution of Σ falling in certain ranges

n	σ_Σ	Proportion falling outside range		
		$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
6	5 32	0·272	0 056	0·0000
7	6·66	0·381	0·030	0·0004
8	8·08	0·275	0·031	0·0004
9	9·59	0·359	0·045	0·0009
10	11·18	0·291	0·047	0·0009

From this β_2 may be obtained and it is evident that as n becomes large β_2 tends to the value 3. In fact it remains below that value, so that the distribution of Σ and therefore of τ is slightly platykurtic. The following table shows the values of β_2 for some values of n. The corresponding values of β_2 for the distribution of ρ are also given and it will be observed that, as judged by β_2, the approach of τ to normality is appreciably quicker than that of ρ.

TABLE III

Values of β_2 in the distribution of Σ and of ρ for certain values of n

n	$\beta_2(\Sigma)$	$\beta_2(\rho)$
5	2·53	2·07
10	2·78	2·54
20	2·89	2·77
30	2·93	2·85

In general, as will be seen below, the moment of order $2s$ is a polynomial of degree n^{3s}.

PROOF OF THE NORMALITY OF τ FOR LARGE n

15. We shall prove that as $n \to \infty$

$$\mu_{2s} \sim \frac{(2s)!}{2^s s!} (\mu_2)^s,$$

where μ_{2s} is the $2s$th moment of the distribution of Σ. In virtue of the symmetry of the distribution moments of odd order vanish and it follows from the Second Limit Theorem of Probability (see Fréchet & Shohat, 1931) that the distribution

of Σ, and hence that of τ, tends to normality in the sense that the frequency between τ_1 and τ_2 tends to

$$\frac{1}{\sigma\sqrt{(2\pi)}}\int_{\tau_1}^{\tau_2} e^{-\frac{x^2}{2\sigma^2}}\,dx.$$

16. Consider the effect of operating on the product f (equation (3)) by $\theta \equiv x\dfrac{\partial}{\partial x}$. The first operation will result in a sum of terms of type

$$\{-rx^{-r}-(r-2)\,x^{-(r-2)}-\ldots+(r-2)\,x^{(r-2)}+rx^{r}\}$$

multiplied by the remaining terms of f unchanged. When x is put equal to unity we may write this as the sum of terms

$$\frac{-r-(r-2)-\ldots+(r-2)+r}{1+1+\ldots+1+1}\,n! = \frac{-r-(r-2)-\ldots+(r-2)+r}{r}\,n!.$$

Similarly the second operation will bring out terms like

$$\frac{r^2+(r-2)^2+\ldots+(r-2)^2+r^2}{r}\,n!$$

and $$n!\left\{\frac{-r-(r-2)-\ldots+(r-2)+r}{r}\right\}\left\{\frac{-t-(t-2)-\ldots+(t-2)+t}{t}\right\}.$$

Generally, operating $2s$ times will bring out terms like

$$n!\left\{\frac{r^{2s}+(r-2)^{2s}+\ldots+(r-2)^{2s}+r^{2s}}{r}\right\},$$

$$n!\left\{\frac{-r^{2s-1}-(r-2)^{2s-1}-\ldots+(r-2)^{2s-1}+r^{2s-1}}{r}\right\}$$
$$\times\left\{\frac{-t-(t-2)-\ldots+(t-2)+t}{t}\right\},$$

etc.

When x is put equal to unity any term beginning with an odd superscript in the powers will vanish. Consider now the sum of terms like

$$n!\left\{\frac{r^2+\ldots+r^2}{r}\right\}\left\{\frac{t^2+\ldots+t^2}{t}\right\}\ldots\left\{\frac{u^2+\ldots+u^2}{u}\right\}, \qquad \ldots\ldots(9)$$

containing s factors.

It will be proved below that this term contributes the greatest power of n to the total sum giving $\mu_0\mu_{2s}$.

Further, in virtue of the multinomial form of Leibniz' theorem, the factor by which this term is multiplied in the expansion of $(\theta^{2s}f)$ is

$$\frac{(2s)!}{2!\,2!\ldots 2!} = \frac{(2s)!}{2^s}.$$

Hence, since $\mu_0 = n!$ we have

$$\mu_{2s} \sim \frac{(2s)!}{2^s}\{\text{sum of terms like (9)}\}. \qquad \ldots\ldots(10)$$

Each term in (9) is of type

$$\frac{1}{r}\{r^2+(r-2)^2+\ldots+(r-2)^2+r^2\},$$

i.e. is of order $\frac{r^2}{3}$. The summation will therefore tend to the sum of terms like $\frac{1}{3^s}\{1^2 . 2^2 . \ldots . s^2\}$, each term containing s squares of the numbers $1, 2, \ldots, (n-1)$. Call this Π_s.

Then Π_s is $\frac{1}{s!}$ times the sum of terms in

$$\frac{1}{3^s}\{1^2+2^2+\ldots+(n-1)^2\}^s, \qquad \ldots\ldots(11)$$

which contain s different factors.

Now (11) is of order $\frac{n^{3s}}{9^s} \sim (\mu_2)^s$. Hence if the product term Π_s tends to the sum (11)

$$\Pi_s \sim \frac{(\mu_2)^s}{s!},$$

and in virtue of (10)

$$\mu_{2s} \sim \frac{(2s)!}{2^s}\frac{(\mu_2)^s}{s!}.$$

To complete the demonstration, we have therefore to show that (11) tends asymptotically to the sum of its terms $s!\,\Pi_s$, i.e. that sums of terms like

$$1^4 . 2^2 . \ldots . (s-1)^2, \quad 1^6 . 2^2 . \ldots . (s-2)^2$$

tend in comparison to zero.

This may be shown inductively.

Consider first of all

$$\{1^2+2^2+\ldots+(n-1)^2\}^2 = 2\Pi_2 + 1^4+2^4+\ldots+(n-1)^4.$$

The expression on the left $\sim \frac{n^6}{9}$. But the sum of fourth powers on the right $\sim \frac{n^5}{5}$, which is of lower order. Hence the sum on the right $\sim 2\Pi_2$. Multiplying by $\{1^2+2^2+\ldots+(n-1)^2\}$ we have

$$\{1^2+2^2+\ldots+(n-1)^2\}^3 \sim 2\Pi_2\{1^2+2^2+\ldots+(n-1)^2\}$$
$$\sim 6\Pi_3 + \text{terms of type } 1^4 2^2.$$

These terms will be less in sum than

$$2\{1^2+2^2+\ldots+(n-1)^2\}\{1^4+2^4+\ldots+(n-1)^4\},$$

which $\sim 2 \cdot \frac{n^3}{3} \cdot \frac{n^5}{5}$, of order 8. But the expression on the left is of order 9. Hence $\{1^2+2^2+\ldots+(n-1)^2\}^3 \sim 6\Pi_3$ and so on.

We can now justify the assertion that the maximum power of n arises from terms like $(1^2 . 2^2 . \ldots . s^2)$. In fact, by a similar line of reasoning to that just given it will be seen that sums of terms of type $\{1^4 . 2^2 . \ldots . (s-1)^2\}$, etc. are of lower order.

The demonstration is complete.

17. It appears therefore that the coefficient τ has a good claim to serious consideration as a measure of rank correlation. It is easily calculable. In the important case of the distribution wherein all possible rankings occur equally frequently its standard error is known; for the values of n likely to be required in practice it may be taken to be normally distributed, and where there is doubt the distribution can be obtained in an exact form.

It should also be remarked that τ has a natural significance. An observer who is given a set of objects (such as coloured discs) to rank appears to follow a process something like this: First of all he searches for the beginning of the series, say the disc of lightest shade. Having selected a disc, he compares it with each of the remainder to verify the propriety of his choice. The coefficient τ gives him one mark for each comparison which is made correctly, and subtracts a mark for each error.* When the first disc is selected, he proceeds as before with a second; and so on. τ follows this process exactly. It appears to be a logical measure of ranking carried out by the process and should therefore prove useful in psychological work.

REFERENCES

FRÉCHET, M. & SHOHAT, J. (1931). "A proof of the generalised second limit theorem in the theory of probability." *Trans. Amer. Math. Soc.* **33**, 533.

HOTELLING & PABST (1936). "Rank correlation and tests of significance involving no assumptions of normality." *Ann. Math. Statist.* **7**, 29.

* Inasmuch as comparisons between extremes in the series will generally be easier than comparisons between neighbouring members it might in some cases be preferable to weight the marking given to different comparisons according to some selected scale. The determination of such a scale, however, would depend to some extent on the circumstances of individual cases and would present considerable difficulty where no objective order is known to exist, apart from adding greatly to the complexity of the distribution of the measure so obtained.

INDIAN RACES IN THE UNITED STATES. A SURVEY OF PREVIOUSLY PUBLISHED CRANIAL MEASUREMENTS*

BY GERHARDT VON BONIN AND G. M. MORANT

1. INTRODUCTION

To the physical anthropologist the American aborigines present some most interesting problems. Are they a homogeneous population or do they show racial divergences similar to those found for the populations of other continents? For how long have they inhabited the New World, and how did they migrate into it originally? Answers to these questions should give us not only a sound knowledge of the American Indians, but should also afford a further insight into human evolution.

The present paper is intended as but a modest contribution towards the solution of such problems. It represents a statistical discussion of the existing craniological material, with the object in view of arriving at a racial classification similar to those already given for the greater part of the Old World and for Australia (Kitson, 1931; Morant, 1925, 1927, 1928; Woo & Morant, 1932). It is further restricted in its scope to the area of the United States. Treatment at the same time of data for Canadian Indian peoples would have been convenient, but in fact there is no suitable material available for them.

Almost a century ago, Morton (1839) published measurements on 147 American Indian skulls of adults. In his preface he remarks that his ample material had enabled him "to give a full exposition of a subject which was long involved in doubt and controversy". Unfortunately, his craniometric technique has now become obsolete. The next notable contribution to the subject was published in 1892 when Virchow provided descriptions of twenty-eight skulls, from different tribes, including artificially deformed specimens. Clearly, his data do not lend themselves to a statistical treatment.

In recent years a much larger amount of material has been published by Hrdlička, by Gifford and by Hooton. It is possible to attempt a racial classification on this basis, although even with as many as 1167 undeformed male skulls it can only be of a provisional nature. Far more evidence would be required for the "full exposition" which Morton had in mind.

* Joint contribution from the Department of Anatomy, University of Illinois, Chicago, and from the Galton Laboratory, University College, London.

2. THE SOURCES OF THE MATERIAL AND THE METHODS OF TREATING IT EMPLOYED

The measurements of crania of Indians of the United States to be discussed were taken from the following publications:

(a) Aleš Hrdlička, "The Anthropology of Florida", *Publications of the Florida State Historical Society*, No. 1 (1922), pp. 140. Eight absolute and eight indicial measurements are given where possible for each skull.

(b) Edward Winslow Gifford, "Californian Anthropometry", *Univ. Calif. Publ. Amer. Archaeol. Ethn.* **22** (1926), 217–390. Individual measurement taken by several anthropologists are given, and the number of characters recorded is not the same for all the series. For the best described skulls fourteen absolute measurements (including the heights and breadths of both orbits) and eight indicial measurements are given. Nothing is said about the techniques followed by the different observers, but these are apparently considered to be identical for all practical purposes, at any rate, and to give results directly comparable with Hrdlička's.

(c) Aleš Hrdlička, "Catalogue of Human Crania in the United States National Museum Collections. The Algonkin and Related Iroquois; Siouan, Caddoan, Salish and Sahaptin, Shoshonean, and Californian Indians", *Proc. U.S. Nat. Mus.* **69**, Art. 5 (1927), 1–127. There are eleven absolute and seven indicial measurements recorded in this part of the *Catalogue*.

(d) Aleš Hrdlička, "Catalogue of Human Crania in the United States National Museum Collections. Pueblos. Southern Utah Basket-Makers. Navaho". *Proc. U.S. Nat. Mus.* **78**, Art 2 (1931), 1–95. The measurements given comprise all those in the 1927 part of the *Catalogue* together with seven other chords, three other indices, two angles and one measurement of the mandible.

(e) Earnest Albert Hooton, *The Indians of Pecos Pueblo. A Study of their Skeletal Remains*, New Haven (1930), pp. xxvii + 391. It is said to be improbable that any of the skeletons reported on are much more than 1000 years old, and they cover a period extending down to the early nineteenth century. Individual measurements are not given, but means and standard deviations are recorded for a number of groups. The characters treated include nearly all in Hrdlička's tables and a number of others which are not available for any other North American series.

No adequate definitions of the measurements recorded are given in any of the above sources, though Dr Hrdlička (1919) has elsewhere described his technique in detail.* It is based on the International (Monaco) Agreement of 1906,

* The following symbols are used to denote measurements in tables below: C = capacity, L = maximum glabella-occipital length, B = maximum calvarial breadth, H' = basio-bregmatic height, LB = chord nasion to basion, GL = chord basion to alveolar point, $G'H$ = chord nasion to alveolar point, J = maximum bizygomatic breadth, NB = maximum breadth of pyriform aperture, NH = chord from nasion to subnasal point, O_1' (R or L) = orbital breadth from dacryon, and O_2 (R or L) = orbital height perpendicular to O_1'. $N\angle$, $A\angle$, and $B\angle$ are the angles of the fundamental triangle of which the apices are the nasion, alveolar point and basion.

but some modifications are introduced. It was to be expected that the measurements of the other American observers were taken by following either Hrdlička's instructions or those of the Monaco scheme. No reason to question this assumption was found except in the case of the orbital breadths recorded in Gifford's paper. It is shown to be extremely probable (see pp. 99–101 below) that these were not obtained in accordance with Hrdlička's definition, and hence the means of the orbital breadths and indices for Gifford's series were omitted in making comparisons with others. Hrdlička's definitions are discussed (Morant, 1937, pp. 2–4), in a paper dealing with his Eskimo material. Considerably more than half of the skulls treated below were measured by him and his assistants. A few of his Californian series may be supposed to represent the same populations as a few of Gifford's, but otherwise there is no duplication of this kind. It is to be regretted that the craniologists cited failed to record a number of customary measurements. There are no arcs available for any of the series except Prof. Hooton's Pecos Pueblo, and this is unfortunately found to be unsuitable for comparative purposes.

Karl Pearson's method of the coefficient of racial likeness is used in the treatment of the material given below, both in considering how suitable series can best be made up, and in estimating the resemblances of the types defined by the series finally selected.* This method has recently been criticized with little regard to the fact that its limitations and imperfections were fully recognized by its inventor, or to the way in which it has been used in practice for more than 10 years. For practical purposes, the crude coefficient of racial likeness remains still the best means to estimate whether two samples may be considered to represent the same population or not, and the reduced coefficient remains an effective criterion of the presence or absence of a racial bond between two differentiated samples. Past experience gives no reason to believe that the method of the coefficient of racial likeness fails to provide close approximations to the results which could be obtained by applying theoretically more correct formulae, such as those taking into account all the intercorrelations of the measurements used but which have the disadvantage of involving many times as much arithmetical labour. In the present case it is unfortunate that the available data are so limited that coefficients can be computed from only 11 to 18 characters instead of from 31 as has been done in the past whenever possible. The desirability of using this large number of characters has repeatedly been pointed out. But the method of the coefficient of racial likeness—being admittedly a "stop-gap"—is not a simple rule of thumb. The way in which its values have to be interpreted in order to yield useful results has to be determined empirically. It is precisely this point that the present paper will throw into relief as we shall see subsequently. In calculating all the coefficients the standard deviations of the long Egyptian *E*

* The formulae used in practice to compute the crude and reduced coefficients are given by Cleaver (1937, pp. 100, 102).

(see Pearson & Davin, 1924) series were used, and it is shown in section 8 that these are remarkably close to the average standard deviations for the American Indian series.

3. COMPARISONS OF SERIES OF MALE CRANIA OF CALIFORNIAN INDIANS

The Californian data will be treated first as they are more adequate than those for any other small group of North American Indians. In the 1927 section of his *Catalogue* (pp. 102–25) Hrdlička gives individual measurements of 200 male Californian crania. They are divided into ten series on a geographical basis, and on account of their small sizes no use has been made of six of these series in the present paper. The means for the remaining four are in Table I below,* and the reduced coefficients of racial likeness between them are in Table II. The localities from which the material was obtained are indicated approximately on the map (Fig. 1).

The two mainland series are clearly differentiated from one another and from the two island series, but these last give a negative coefficient. The fact that they cannot be distinguished is not surprising, as the islands are only five miles apart. The identity in type of the two populations cannot be considered well established, however, in view of the small size of the Santa Rosa sample. The means for the combined series from the two islands were calculated and these lead to the reduced coefficients given in the last column of Table II. The neighbouring island and mainland (Santa Barbara County) series are seen to be very similar in type, while the San Francisco series bears a closer resemblance to the Santa Barbara than to the island series. There is thus a correspondence between the resemblances of these types and the geographical positions of the populations they represent.

In his 1926 paper Gifford gives individual measurements taken by himself and other anthropologists of series of Californian Indian crania preserved in several museums, and these are all different from the specimens which appear in Hrdlička's *Catalogue*. It is presumed that the definitions used by all the observers accord with Hrdlička's, and comparisons between his readings and those of Kroeber and Sand on the same twenty specimens are given. These only relate to 9 characters, not including orbital measurements, and a fairly satisfactory agreement is indicated.

Gifford's material is divided up into a large number of small series representing subdivisions of counties and covering the whole of the state. In the majority of cases grouping of these is necessary in order to obtain samples large enough for statistical purposes. The following arrangement was adopted:

(*a*) Northern counties: Gifford's areas 1*b*, 2*a*, 3, 6*e*, 6*f*, 7*d*, 16*a*, 16*b*, 16*c*, 17*a*, 17*b*, 17*c*, 18*a*, 18*c* and 18*d*.

(*b*) Costanoan people: Gifford's areas 19*a*, 19*b*, 19*c* and 19*f*. This is roughly

* The four series in question are those indicated as measured by Hrdlička alone.

TABLE I

*Mean measurements of male series of Californian crania measured by Hrdlička (H.) and Gifford (G.)**

Series	Series used initially						Series used finally					
	San Francisco Bay and vicinity	Costanoan: vicinity of San Francisco Bay	Santa Cruz Island	Santa Rosa Island	Santa Cruz Island	Santa Rosa Island	Northern California	Central California (Yokuts)‡	San Francisco Bay and vicinity (including Costanoan)	Santa Barbara County (mainland)	Santa Cruz and Santa Rosa Islands	Santa Catalina, San Clemente and San Nicolas Islands
Measured by ..	H	G.	H	H.	G.	G.	G.	G.	H. and G.	H.	H. and G.	G.
C	1372·4 (22)	—	1393·7 (57)	1409 4 (18)	1280·7 (53)	—	—	—	1372·4 (22)	1389·7 (27)	1349·1 (128)	1403·9 (26)
L	182 3 (31)	183 4 (115)	181·5 (66)	179·0 (20)	178·0 (55)	181·0 (54)	178 7 (54)	181·2 (45)	183·2 (146)	179·4 (48)	180·1 (195)	190 0 (40)
B	140·1 (31)	138 8 (112)	140 7 (65)	140·8 (18)	138·4 (54)	141·1 (55)	141·6 (54)	144·3 (44)	139 1 (143)	138 3 (45)	140 2 (192)	136·7 (41)
H'	136·6 (27)	136·1 (76)	130·6 (64)	130·2 (20)	127·4 (53)	128 8 (51)	134·8 (51)	137·2 (42)	136·2 (103)	131·7 (45)	129·2 (188)	128·4 (35)
J	136·7 (18)	135·5 (55)	136 5 (59)	135 7 (16)	133·7 (50)	135 1 (47)	136·65 (40)	140 9 (32)	135 8 (73)	135·4 (28)	135·2 (172)	137·6 (38)
LB	—	100·1 (75)	—	—	96 8 (51)	98·5 (51)	99·5 (49)	100 5 (41)	100·1 (75)	—	97 65 (102)	100·8 (34)
GL	—	99·6 (54)	—	—	97·6 (49)	99·2 (49)	98·7 (42)	99·5 (39)	99·6 (54)	—	98·4 (98)	98·5 (31)
$G'H$	73·0 (22)	70·2 (71)	71·2 (67)	70·9 (20)	70·0 (50)	69 9 (54)	70·2 (44)	73·7 (44)	70·9 (93)	69 6 (47)	70·4 (191)	74·3 (36)
NB	24·6 (26)	24·5 (75)	23·9 (66)	23·6 (22)	23 6 (55)	23·1 (56)	25·0 (48)	25 0 (44)	24·5 (101)	23·8 (51)	23 5 (199)	25 45 (40)
NH	51·0 (26)	49 8 (75)	50·4 (66)	49·9 (22)	49·3 (55)	50 8 (56)	49 6 (48)	51 0 (45)	50·1 (101)	49·1 (49)	50·2 (199)	53·2 (40)
O_2R†	34 6 (27)	—	35·2 (65)	35·2 (21)	35·0 (55)	—	—	—	34·6 (27)	34·5 (45)	35 1 (141)	37·3 (25)
O_2L†		—			35·1 (55)	—	—	—				37·3 (25)
$O_1'R$†	38·4 (27)	—	38·3 (65)	38·5 (21)	[40·2 (55)]‖	—	—	—	38 4 (27)	38·3 (44)	38·3 (86)§	[42 4 (25)]‖
$O_1'L$†		—			[39 9 (55)]‖	—	—	—				[41·8 (25)]‖
100 B/L	76·9 (31)	75·7 (108)	77·6 (65)	78·9 (18)	77·9 (54)	77·9 (54)	79·3 (53)	79 6 (42)	76 0 (139)	77·2 (44)	77·9 (191)	71·9 (40)
100 H'/L	{74·9 (27)}	74·8 (74)	{72·0 (64)}	{72·7 (18)}	71·6 (53)	71 5 (51)	75·6 (50)	75·8 (40)	{74·8 (103)}	{73·4 (45)}	{71·7 (188)}	67 6 (34)
100 B/H'	{102 6 (27)}	{102·0 (76)}	{107·7 (64)}	{108 1 (18)}	{108 6 (53)}	{109·5 (51)}	{105 0 (51)}	{104·2 (42)}	{102 1 (103)}	{105 0 (45)}	{108 5 (188)}	{106·5 (35)}
100 NB/NH	48·5 (26)	49·1 (72)	47·5 (66)	47·4 (22)	47·9 (55)	45·6 (56)	50·5 (47)	49 2 (44)	48 9 (98)	48·4 (48)	47 1 (199)	48 0 (40)
100 O_2/O_1'†	90 2 (27)	—	92·0 (65)	91·6 (21)	[87·7 (55)]‖	—	—	—	90·2 (27)	89 6 (44)	91·9 (86)§	[88·6 (25)]‖
$N\angle$	—	{68°·6 (54)}	—	—	{69°·4 (49)}	{70°·0 (49)}	{68°·9 (42)}	{67°·6 (39)}	{68°·6 (54)}	—	{69°·7 (98)}	{66°·2 (31)}
$A\angle$	—	{69°·9 (54)}	—	—	{68°·3 (49)}	{68°·5 (49)}	{69°·6 (42)}	{68°·9 (39)}	{69°·9 (54)}	—	{68°·4 (98)}	{70°·0 (31)}
$B\angle$	—	{41°·5 (54)}	—	—	{42°·3 (49)}	{41°·7 (49)}	{41°·5 (42)}	{43°·5 (39)}	{41°·5 (54)}	—	{41°·9 (98)}	{43°·8 (31)}

* The indices and angles in curled brackets were found from the means of the chords involved instead of from values for individual skulls.

† Hrdlička's orbital heights and breadths and the orbital indices for all the series were obtained from average values for the right and left sides.

‡ Omitting one skull (No. 12–1668) for which the L given is 148: this is probably an error, or if correct the specimen is pathological.

§ For Hrdlička's measurements only.

‖ These values were not used in computing coefficients of racial likeness: see pp. 99–101 of text

TABLE II

*Reduced coefficients of racial likeness for male series of Californian crania measured by Hrdlička**

	San Francisco Bay and vicinity	Santa Barbara County	Santa Rosa Island	Santa Cruz Island	Santa Rosa and Santa Cruz Islands
$\bar{n}$	26·3	43·7	19·7	64·3	84 1
San Francisco Bay and vicinity	—	8·25 ± 0·75	8·56 ± 1·09	12·94 ± 0·66	13·96 ± 0·61
Santa Barbara County	8·25 ± 0·75	—	3 38 ± 0·91	4·57 ± 0 47	4 43 ± 0 43
Santa Rosa Island	8·56 ± 1·09	3·38 ± 0·91	—	− 1·26 ± 0·82†	—
Santa Cruz Island	12·94 ± 0·66	4·57 ± 0·47	− 1·26 ± 0 82†	—	—

* All the coefficients in this table are based on the same fifteen characters (see Table I).
† The crude coefficient corresponding to this is − 0 38 ± 0·25.

the same area as that from which Hrdlička's "San Francisco Bay and vicinity" series was obtained.

(*c*) Yokut people: Gifford's areas 20*a*, 20*b*, 20*e* and 20*g*; Central California.

(*d*) Santa Rosa Island.

(*e*) Santa Cruz Island.

(*f*) Santa Catalina, San Clemente and San Nicolas Islands.

Male means for series made up in this way are given in Table I.* Gifford also gives measurements for other small series, but these come from scattered localities, and it was felt that pooling of them to give sufficiently large samples would not be justified.

Comparisons may be made first between the pairs of series made up from Hrdlička's and Gifford's data which may be supposed to represent the same populations. The "San Francisco Bay and vicinity" series of the former corresponds with the Costanoan of the latter, and the crude coefficient of racial likeness between them for 11 characters is 0·87 ± 0·29. It should be noted that no comparisons of orbital measurements are included as no means for these are available for the Costanoan series. The absence of any evidence of differentiation is very satisfactory, and there can be no objection to pooling the two to form a longer series. The other corresponding groups relate to Santa Rosa and Santa Cruz Islands. A comparison of the means for these shows at once that the orbital breadths ($O_1' R$ and $O_1' L$) and index for Gifford's Santa Cruz series differ very significantly from those for Hrdlička's Santa Cruz and Santa Rosa series. This obviously suggests that two different definitions were followed in finding the

* They are the six indicated as measured by Gifford alone.

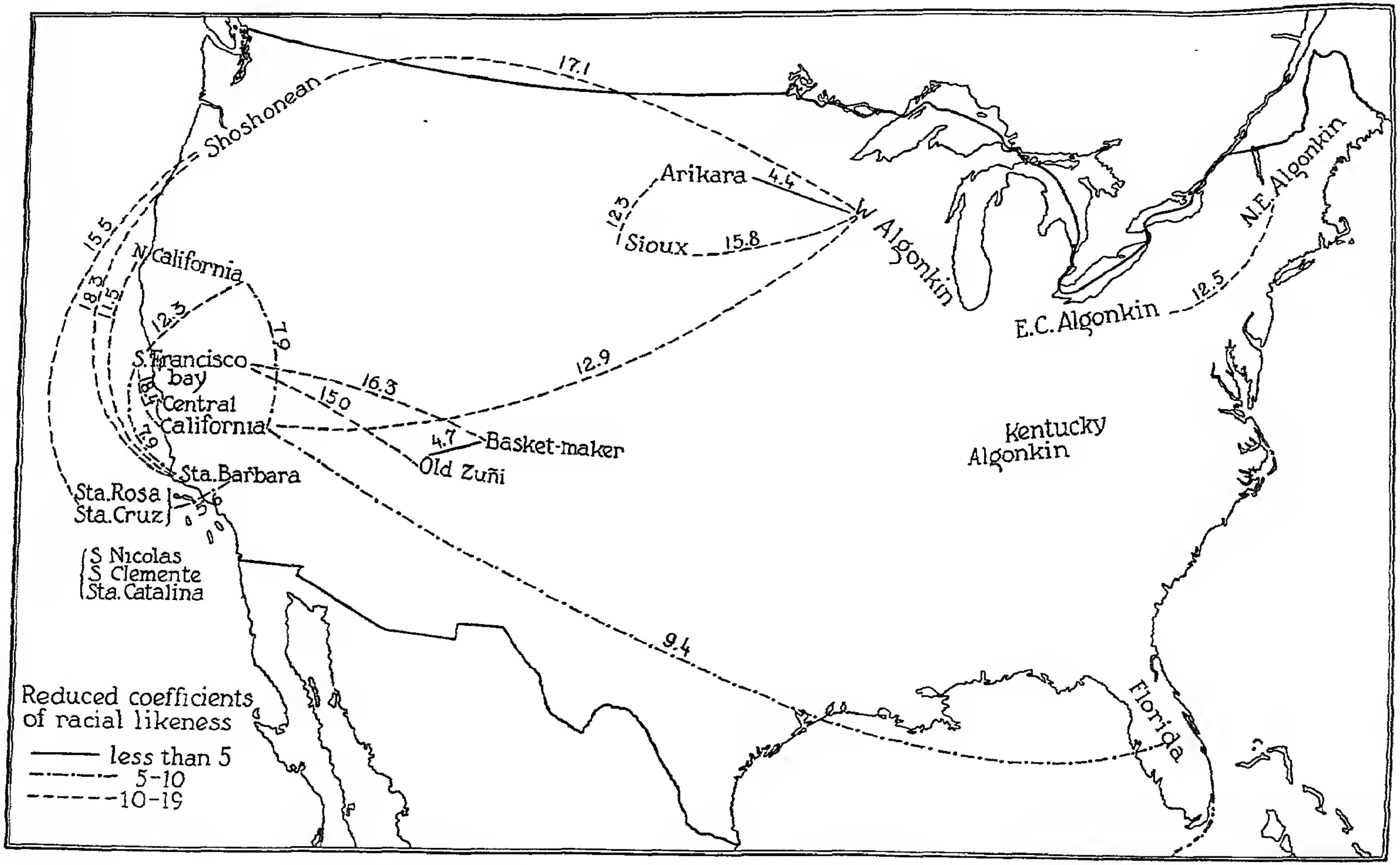


Fig. 1. Approximate locations of American Indian Cranial Series and Relationships between them.

orbital breadths. No comparison of this character could be made in the case of the San Francisco series, but it may be noted that the means for Gifford's Santa Catalina, San Clemente and San Nicolas Islands series are also markedly greater than all Hrdlička's. The latter defines his to be the dacryal breadth and his values were accepted, while all orbital breadths and indices given by Gifford were omitted in computing coefficients of racial likeness. These constants for the four island series are given in Table III. Hrdlička's Santa Rosa series is by far the shortest of the four and all the lowest coefficients are found with it. Two of them differ

TABLE III

Crude coefficients of racial likeness for series of Californian crania from Santa Rosa and Santa Cruz Islands

	Santa Rosa* (Hrdlička)	Santa Cruz† (Hrdlička)	Santa Rosa‡ (Gifford)	Santa Cruz§ (Gifford)
Santa Rosa (Hrdlička)*	—	-0.38 ± 0.25 (15)	0.43 ± 0.29 (11)	1.58 ± 0.26 (13)
Santa Cruz (Hrdlička)†	-0.38 ± 0.25 (15)	—	1.71 ± 0.29 (11)	4.98 ± 0.26 (13)
Santa Rosa (Gifford)‡	0.43 ± 0.29 (11)	1.71 ± 0.29 (11)	—	2.35 ± 0.25 (14)
Santa Cruz (Gifford)§	1.58 ± 0.26 (13)	4.98 ± 0.26 (13)	2.35 ± 0.25 (14)	—

* $\bar{n} = 19.7$ for 15 characters and 19·5 for 11 and 13.
† $\bar{n} = 64.3$ for 15 characters, 64·7 for 11 and 64·2 for 13.
‡ $\bar{n} = 53.2$ for 11 characters and 52·4 for 14.
§ $\bar{n} = 53.5$ for 13 characters and 52·6 for 14.

insignificantly from zero, while the third is significantly different from zero. The remaining three series are approximately equal in length, so their coefficients with one another may be supposed to measure the resemblances of the types, as in the case of reduced coefficients of racial likeness. Some curious relationships are found. Hrdlička's Santa Cruz and Gifford's Santa Rosa series are seen to resemble one another more closely than either resembles Gifford's Santa Cruz series. An examination of the mean measurements in Table I shows that the type of the last is so distinguished on account of its smaller size. Still omitting the orbital breadth, every one of its means for absolute measurements is less than the corresponding value for Hrdlička's series from the same island. The divergence of the two Santa Cruz series is thus probably due to sexing and not to differences in the ways the measurements were taken, or to the fact that two different populations are represented. For Hrdlička's combined series from Santa Cruz and Santa Rosa Islands and Gifford's combined series from the same two islands a crude coefficient of 4.50 ± 0.26 is found for 13 characters. All the means of the absolute measurements for the former exceed those for the latter, and by far the most significant difference is for the capacity ($\alpha = 32.9$). In all these comparisons no significant differences are found for the indices. It is certainly curious that Gifford's Santa Cruz series should be distinguished by the smaller size of its type,

not only from both Hrdlička's but also from Gifford's Santa Rosa series. The hypothesis that differences in sexing are responsible for these relationships seems to be a plausible one, and accordingly all four series were pooled for comparisons with others in the hope that the resulting means (given in the penultimate column of Table I) give as fair a representation of the male type of a homogeneous population as any which could be obtained from the data available.

Having carried out the grouping described, six series of male Californian Indian crania were made up from the measurements provided by Hrdlička and Gifford. The means for these are given in Table I and the reduced coefficients of racial likeness between them in Table IV. It should be noted that these last constants are based on differing numbers of characters ranging from 11 to 18. In all cases as many of the 31 characters (used when possible in calculating the coefficients) as are available were employed, and it is to be regretted that some of the numbers fall far short of this total. The difference between a reduced coefficient based on 20 characters and another for the same two sets of means based on 30 characters is generally found to be of little account, but a change from 11 to 18 characters is likely to be of far more consequence. The reduced coefficients in Table V were calculated with the object of gaining some idea of the effects that such a change may have. They are between the pooled series from Santa Cruz and Santa Rosa Islands and the five others. The second column gives values calculated only for the 11 characters common to all six series, and the third for those obtained when all possible characters are used in each case. The latter values are all less than the corresponding former ones, owing to the fact that the characters added tend to show less significant differences than the others, on the average. The reduction is only marked in one case, however, and the two sets arrange the five series in the same order. The use of differing numbers of characters for different coefficients is far from satisfactory, but if comparisons are to be made with data for other continental areas, it appears better to use all of the selected list available in each case, rather than a constant number considerably smaller than those which it has generally been possible to use.

The connections between the series provided by the lowest orders of reduced coefficients of racial likeness are shown in Fig. 1. There appears to be a fairly close association between the relationships of the types and their geographical positions in the case of five of the series, but the remaining one—from Santa Catalina, San Clemente and San Nicolas Islands—is widely removed from the others. A comparison of the means shows at once that the last type has high coefficients with all the others chiefly on account of its greater calvarial length and lower cephalic and height-length indices. Even if it be excluded, the Californian types show greater diversity than is generally found for adjoining populations inhabiting a small region. In particular the neighbouring Costanoan and Yokut groups are far less similar than might have been expected

Hrdlička, who had not measured any material from the southern islands,

TABLE IV

*Reduced coefficients of racial likeness for male series of Californian crania measured by Hrdlička (H.) and Gifford (G.)**

	Northern California	Central California (Yokuts)	San Francisco Bay and vicinity (including Costanoan)
Measured by	G.	G.	H. and G.
$\bar{n}$	48·1†	41·6‡	82·7§
Northern California	—	7·89±0·57 (14)	12 35±0·39 (14)
Central California	7·89±0·57 (14)	—	18 43±0·44 (14)
San Francisco Bay and vicinity	12·35±0 39 (14)	18·43±0·44 (14)	—
Santa Barbara County	11·47±0·61 (11)	31·37±0 66 (11)	7·85±0·42 (15)
Santa Cruz and Santa Rosa Islands	21 30±0·34 (14)	37·15±0 38 (14)	25 00±0·21 (18)
Santa Catalina, San Clemente and San Nicolas Islands	92·08±0·61 (14)	94·03±0·65 (14)	52·39±0·47 (16)

	Santa Barbara County	Santa Cruz and Santa Rosa Islands	Santa Catalina, San Clemente and San Nicolas Islands
Measured by	H.	H. and G.	G.
$\bar{n}$	43·7‖	157·8¶	35·4**
Northern California	11·47±0·61 (11)	21 30±0·34 (14)	92·08±0·61 (14)
Central California	31·37±0·66 (11)	37·15±0·38 (14)	94·03±0·65 (14)
San Francisco Bay and vicinity	7·85±0·42 (15)	25·00±0 21 (18)	52·39±0·47 (16)
Santa Barbara County	—	5·56±0·35 (15)	70·10±0 67 (13)
Santa Cruz and Santa Rosa Islands	5·56±0·35 (15)	—	56·53±0·41 (16)
Santa Catalina, San Clemente and San Nicolas Islands	70·10±0·67 (13)	56 53±0·41 (16)	—

* The $\bar{n}$'s are the mean numbers of skulls for all the coefficients of racial likeness characters available. For some comparisons it is not possible to use all these characters which are available and the $\bar{n}$'s in these cases are given in the footnotes below.

† $\bar{n}=49{\cdot}1$ for 11 characters, omitting LB, $N\angle$ and $A\angle$.

‡ $\bar{n}=42{\cdot}2$ for 11 characters, omitting LB, $N\angle$ and $A\angle$.

§ $\bar{n}=99{\cdot}0$ for 14 characters, omitting C, O_1', O_2 and $100\ O_2/O_1'$, 87·1 for 15 characters, omitting LB, $N\angle$ and $A\angle$; 89·7 for 16 characters, omitting O_1' and $100\ O_2/O_1'$.

‖ $\bar{n}=45{\cdot}0$ for 11 characters, omitting C, O_1', O_2 and $100\ O_2/O_1'$; 43·6 for 13 characters, omitting O_1' and $100\ O_2/O_1'$.

¶ $\bar{n}=171{\cdot}4$ for 14 characters, omitting C, O_1', O_2 and $100\ O_2/O_1'$; 169·5 for 15 characters, omitting LB, $N\angle$ and $A\angle$; 166·8 for 16 characters, omitting O_1' and $100\ O_2/O_1'$.

** $\bar{n}=36{\cdot}2$ for 13 characters, omitting LB, $N\angle$ and $A\angle$; 36·8 for 14 characters, omitting C and O_2.

TABLE V

Reduced coefficients of racial likeness based on different sets of characters: male series of Californian crania

Santa Cruz and Santa Rosa Islands with ...	Reduced coefficients for		Additional characters
	11 characters*	All available characters	
Northern California	$24{\cdot}97 \pm 0{\cdot}37$	$21{\cdot}30 \pm 0{\cdot}34$ (14)	LB, $N\angle$, $A\angle$
Central California	$37{\cdot}76 \pm 0{\cdot}42$	$37{\cdot}15 \pm 0{\cdot}38$ (14)	LB, $N\angle$, $A\angle$
San Francisco Bay and vicinity	$37{\cdot}47 \pm 0{\cdot}21$	$25{\cdot}00 \pm 0{\cdot}21$ (18)	C, LB, O_1', O_2, $N\angle$, $A\angle$, $100\ O_2/O_1'$
Santa Barbara County	$6{\cdot}20 \pm 0{\cdot}40$	$5{\cdot}56 \pm 0{\cdot}35$ (15)	C, O_1', O_2, $100\ O_2/O_1'$
Santa Catalina, San Clemente and San Nicolas Islands	$64{\cdot}78 \pm 0{\cdot}45$	$56{\cdot}53 \pm 0{\cdot}41$ (16)	C, LB, O_2, $N\angle$, $A\angle$

* Viz. L, B, H', J, NH, NB, $G'H$, $100\ B/L$, $100\ H'/L$, $100\ B/H'$, $100\ NB/NH$.

concluded that "the material from California shows considerable uniformity". Gifford distinguished three main living types, one of which was divided into three subtypes, and seven cranial types. The skull measurements do not appear to justify such an arrangement. The one adopted in this paper has a geographical basis, and it should be pointed out that far more adequate material would be required to delimit accurately the groups distinguished in such a way. The groups used here are obviously of provisional value only.

4. Comparisons of series of male crania of Algonkin and related Indians

Another group of material which can be conveniently considered by itself is provided by the Algonkin and Iroquois series for which measurements are provided by Hrdlička in the 1927 section of his *Catalogue*. Most of these series are too short to be treated singly, and a grouping of them on a geographical basis in order to obtain large enough samples was hence necessitated. The pooling which was first carried out can be seen from the headings of the columns in the upper part of Table VI. The series I*a*, I*b* and I*c* represent adjoining regions in the extreme north-east of the country. The crude coefficients of racial likeness between them for 14 characters (omitting the capacity) are: I*a* and I*b*, $0{\cdot}70 \pm 0{\cdot}25$; I*a* and I*c*, $0{\cdot}48 \pm 0{\cdot}25$; I*b* and I*c*, $1{\cdot}22 \pm 0{\cdot}25$. Only the last of these can be supposed significant, and as it still indicates very close resemblance, it was felt that the pooling of the three series was advisable. The Iroquois series is only distinguished from the other two by the fact that its mean nasal breadth is significantly greater

TABLE VI

Mean measurements of male cranial series referring to Algonkin and related Indian tribes

States (tribes)	Maine, Huron, Massachusetts, Connecticut, Rhode Island	North-west of New York (Iroquois)	New York, Manhattan Island, Long Island, Staten Island	New Jersey (Delaware), Pennsylvania, Maryland, Virginia	Ohio, Indiana, Michigan, Illinois
Group	I*a*	I*b*	I*c*	II*a*	II*b*
C	1558·3 (3)	—	1524·4 (8)	1529·4 (8)	1490·6 (25)
L	188·0 (45)	188·6 (33)	190·5 (42)	185·4 (48)	183·3 (46)
B	137·7 (45)	137·7 (33)	139·5 (42)	139·7 (48)	138·6 (45)
H'	137·9 (41)	138·9 (31)	140·4 (38)	141·5 (30)	141·6 (34)
$G'H$	75·2 (22)	74·8 (21)	73·8 (27)	72·8 (12)	74·8 (24)
J	137·5 (26)	138·4 (23)	138·8 (28)	139 9 (12)	140·3 (19)
NB	25·6 (31)	27·4 (26)	25·6 (32)	27·1 (17)	25·9 (35)
NH	52·3 (31)	53·5 (27)	52·3 (32)	52·7 (18)	53·8 (33)
O_2*	34·4 (33)	33·9 (25)	33·6 (29)	33·9 (22)	34·9 (29)
O_1'*	39·3 (33)	39·0 (23)	39·4 (29)	38·7 (22)	39·6 (29)
100 B/L	73·2 (45)	73·0 (33)	73·3 (42)	75·4 (48)	75·7 (45)
100 H'/L	{73·4 (41)}†	{73·6 (31)}	{73·7 (38)}	{76·3 (30)}	{77·3 (34)}
100 B/H'	{99·9 (41)}	{99·1 (31)}	{99·4 (38)}	{98 7 (30)}	{97·9 (34)}
100 NB/NH	49·5 (31)	51·5 (26)	49·1 (32)	51·3 (17)	48·5 (33)
100 O_2/O_1'*	87·5 (33)	86·8 (23)	85·4 (29)	87·7 (22)	88·1 (29)

States (tribes)	Kentucky	Western; Wisc., Iowa, Miss., Montana, (Cheyenne),‡ (Chippewa),§ (Piegan)‖	North-Eastern	East-Central
Group	III	IV¶	I*a*+I*b*+I*c*	II*a*+II*b*
C	1432·5 (24)	1514·0 (41)	1533·6 (11)	1500·0 (33)
L	177·0 (34)	183·9 (49)	189·0 (120)	184·4 (94)
B	135·8 (34)	142·4 (49)	138·3 (120)	139·2 (93)
H'	139·5 (27)	135 0 (47)	139·0 (110)	141·6 (64)
$G'H$	70·4 (25)	72·9 (41)	74·5 (70)	74·1 (36)
J	136·0 (21)	141·7 (35)	138·2 (77)	140·1 (31)
NB	23·8 (26)	26·2 (46)	26·1 (89)	26·3 (52)
NH	50·9 (29)	53·6 (47)	52·7 (90)	53 4 (51)
O_2*	32·6 (30)	35·2 (39)	34·0 (87)	34·5 (51)
O_1'*	38·1 (27)	39·6 (39)	39·3 (85)	39·2 (51)
100 B/L	76·7 (34)	77·5 (49)	73·2 (120)	75·5 (93)
100 H'/L	78·8 (27)	73·4 (47)	73·6 (110)	76·8 (64)
100 B/H'	97·3 (27)	105·5 (47)	99·5 (110)	98·3 (64)
100 NB/NH	46·8 (26)	49·0 (46)	49·9 (89)	49·5 (50)
100 O_2/O_1'*	85·0 (27)	89·0 (39)	86·6 (85)	87·9 (51)

* The orbital measurements given for individual crania are the averages of the readings for the right and left sides.

† The mean indices in curled brackets were found from the means of the component lengths instead of from indices or individual crania.

‡ From Kansas, Wyoming, Colorado and Nebraska.

§ From North Dakota and Michigan.

‖ From Montana.

¶ Omitting three deformed crania from Iowa.

The pooled means for the three series are given in the lower part of the table, and they will be said to relate to the Algonkin· North-Eastern States series, although the Iroquois do not belong to the Algonkin speaking peoples.

The series II*a* represents States immediately to the south and on the eastern seaboard, and II*b* relates to four larger States to the west. A series from Kentucky was excluded from the latter group because it obviously defines a different type. The area covered by II*a* and II*b* together is much larger than that of the north-eastern States. The two series give a crude coefficient of racial likeness of $1{\cdot}07 \pm 0{\cdot}25$ for the 14 characters, and no single character shows a significant difference as the highest α found is 6·0. The coefficient differs significantly from zero, but it indicates a very close resemblance. Accordingly, the series II*a* and II*b* were combined and the pooled means (in Table VI) will be referred to as those of the Algonkin: East-Central States series. The Kentucky series was treated by itself, and the remaining Algonkin skulls for which measurements are given in Hrdlička's 1927 *Catalogue* were pooled to form the Western Algonkin series. These last were obtained in ten States covering an area greater than that of all the Algonkin States to the east of them put together. This pooled series is still a small sample, and it is obvious that far more abundant material would be required to delimit different regional types of population found among the Algonkin speaking peoples. The partitioning of them adopted here is a *pis aller*, and, again, it can only be considered to be of provisional value.

Means for the four series finally adopted are given in the lower part of Table VI. The Kentucky series is almost too short to use for statistical purposes, but the other three are of fairly adequate lengths. The coefficients of racial likeness between them are given in Table VII. It is surprising to find that there is not a single one indicating a close resemblance. The lowest is for the adjoining groups representing the North-Eastern and East-Central States, but this indicates a greater divergence than that usually found between two neighbouring populations. The aberrance of the Kentucky series is particularly striking, and this is evidently due to the small size of its type. For all the absolute measurements in Table VI except H' the Kentucky series has by far the smallest mean, though all its indices differ insignificantly from those for the series representing the East-Central States. Our conclusions accord with Hrdlička's so far as the statement that the Iroquois "are radically of the same physical type with the Algonkins and cannot be separated from the latter", but his contention that "the extensive Algonkin strains shows almost throughout a clear and distinct physical character" is not confirmed.

Table VIII gives the reduced coefficients of racial likeness between the four Algonkin series, on the one hand, and the six Californian, on the other. All the values are high except one, and it must be remembered that little importance should be attached to even marked differences between high coefficients. The exception is for the comparison between the Algonkin series from the Western

TABLE VII

*Coefficients of racial likeness between male cranial series referring to Algonkin and related Indian tribes**

	$\bar{n}$	North-East States	East-Central States	Kentucky	Western States
		Crude coefficients			
North-East States	91·5	—	8·89 ± 0·25	22·41 ± 0·25	16·90 ± 0·25
East-Central States	58·5	8·89 ± 0 25	—	11·57 ± 0·25	11·93 ± 0·25
Kentucky	27·9	22·41 ± 0 25	11·57 ± 0 25	—	22·92 ± 0·25
Western States	44·1	16·90 ± 0·25	11 93 ± 0·25	22·92 ± 0·25	—
		Reduced coefficients			
North-East States		—	12·46 ± 0 34	52·41 ± 0·58	28·40 ± 0·41
East-Central States		12·46 ± 0·34	—	30 64 ± 0·65	23·74 ± 0 49
Kentucky		52·41 ± 0·58	30·64 ± 0 65	—	67·04 ± 0·72
Western States		28·40 ± 0 41	23 74 ± 0·49	67 04 ± 0 72	—

* All the coefficients in this table are based on the 15 characters for which means are given in Table VI. The $\bar{n}$'s for the series are the mean numbers of skulls on which these means are based.

States and the Central Californian series. A much closer resemblance is indicated in this case than those between the former and the other Algonkin types.

5. Comparisons of other United States series

Mean measurements (given in Table IX) were calculated for the three following groups from data given for male skulls in the 1927 section of Hrdlička's *Catalogue*:

(*a*) *Sioux proper*: Miscellaneous 17, Teton 4, Brulé 15, Oglala 14, Sisseton 4, Yankton 5 and Montana 4. The pooling of this material is necessitated by the fact that the series for single tribes are all too short to make their treatment singly profitable. It is said that they are all closely related in physical type, and the mean measurements for the short series confirm this as far as can be seen. They are all characterized by a low basio-bregmatic height. The region represented is made up by the six most westerly states of the large region from which the skulls making up the Western Algonkin series were obtained.

(*b*) *Arikara*, a Siouan tribe said to be nearly related to the Sioux proper, 54. This is a long enough series to be treated by itself. The skulls composing it were all obtained in South Dakota and some of the Sioux proper and Western Algonkin specimens came from the same State. The other series representing Siouan tribes, and the following ones relating to Caddoan, Salish and Sahaptin tribes, are all too short to be used.

TABLE VIII

*Reduced coefficients of racial likeness for Algonkin and Californian male cranial series**

			Californian Series		
			Northern	Central (Yokuts)	San Francisco Bay and vicinity (including Costanoan)
		$\bar{n}$	49·1	42·2	87·1
Algonkin Series	North-East States	91·5	68·30 ± 0·44 (11)	56·03 ± 0·48 (11)	21·29 ± 0·28 (15)
	East-Central States	58·5	51·03 ± 0·52 (11)	36·01 ± 0·57 (11)	21·22 ± 0·35 (15)
	Kentucky	27·9	40·95 ± 0·80 (11)	49·39 ± 0·85 (11)	29·78 ± 0·58 (15)
	Western States	44·1	25·04 ± 0·61 (11)	12·93 ± 0·66 (11)	22·55 ± 0·42 (15)

			Californian Series		
			Santa Barbara County	Santa Cruz and Santa Rosa Islands	Santa Catalina, San Clemente and San Nicolas Islands
		$\bar{n}$	43·7	169·5	36·2
Algonkin Series	North-East State	91·5	52 80 ± 0·42 (15)	78·03 ± 0·21 (15)	55·40 ± 0·51 (13)
	East-Central States	58 5	52·64 ± 0·49 (15)	81·76 ± 0·28 (15)	100·36 ± 0·59 (13)
	Kentucky	27·9	37·40 ± 0·72 (15)	71·88 ± 0·51 (15)	152·98 ± 0·84 (13)
	Western States	44·1	31·66 ± 0·56 (15)	38·26 ± 0·35 (15)	59·35 ± 0·66 (13)

* The characters on which the coefficients in this table are based can be seen from a comparison of the means available for the different series which are given in Tables I and VI, all the characters in the latter table being used when possible. The $\bar{n}$'s given are for all 15 characters in the case of the four Algonkin and three of the Californian series, for 11 characters in the case of two others and for 13 characters in the case of one other Californian series. In the comparison of the Algonkin with these last three series the Algonkin $\bar{n}$'s are slightly different from the values given for all 15 characters.

(*c*) *Shoshonean:* Bannock 1, miscellaneous unidentified tribes 8 (omitting 5 deformed), Utes and Gosh-Utes 9, and Paiutes and Pah-Vants 6. The tribes of this group are said to form a fairly uniform type. (It should be noted that the female Blackfoot specimen and the Piegan series are included in the Shoshonean section of the *Catalogue* in error. The latter forms part of our Western Algonkin series.) The Shoshonean skulls were obtained in Colorado, from which a few of the Sioux proper and Western Algonkin skulls were obtained, and also from four States to the west unrepresented by any other material dealt with above.

TABLE IX

*Mean measurements of male series of Indian skulls from Western and Southern States**

	Sioux	Arikara	Shoshonean	Basket-maker	Old Zuñi	Pecos Pueblo	
States	North and South Dakota, Nebraska, Wyoming, Colorado, Kansas and Montana	South Dakota	Idaho. Oregon, Nevada, Utah and Colorado	South-East Utah	New Mexico	New Mexico	Florida
C	1486·2 (57)	1485 3 (43)	1378·7 (20)	1341·8 (31)	1306·5 (26)	1338·7 (31)	—
L	185·5 (63)	182·8 (51)	182·8 (24)	180 7 (33)	175·8 (34)	175·7 (46)	179·9 (121)
B	143·7 (63)	142·5 (51)	138·7 (23)	131·9 (33)	133·4 (35)	137·8 (45)	144·6 (121)
H′	130·1 (61)	134·6 (50)	128·5 (21)	133·3 (30)	133·5 (30)	137·1 (34)	141·4 (87)
LB	—	—	—	98·9 (33)	99 5 (27)	102·7 (27)	—
GL	—	—	—	96·7 (29)	98·0 (23)	98·6 (25)	—
G′H	76·0 (56)	76·0 (51)	72·3 (23)	73·8 (33)	73·3 (32)	72·85 (112)	74·1 (65)
J	143·0 (59)	141·5 (51)	139·2 (24)	134 8 (36)	134·5 (31)	138·6 (102)	140·3 (65)
NB	26·8 (60)	26·2 (53)	25·5 (23)	25·2 (40)	25·2 (32)	25·8 (126)	25·1 (73)
NH	55·1 (61)	55·6 (53)	52·3 (23)	51·3 (41)	51·0 (32)	51·0 (125)	52·6 (75)
O_2R	36·2 (57)	36·4 (43)	34·8 (22)	34·7 (40)	34·8 (30)	34 8 (119)	—
O_2L					34·9 (29)	34·9 (117)	—
$O_1'R$	40·0 (57)	39·8 (43)	39·5 (22)	38·0 (39)	38·0 (30)	39 9 (119)	—
$O_1'L$					37·8 (29)	39·5 (117)	—
100 *B/L*	77·5 (63)	78 0 (51)	75·9 (23)	73·0 (33)	75 9 (34)	78 3 (43)	80·4 (121)
100 *H′/L*	{70·1 (61)}	{73 6 (50)}	{70·3 (21)}	{73·8 (30)}	{75·9 (30)}	78·1 (34)	78·8 (87)
100 *B/H′*	{110·5 (61)}	{105·9 (50)}	{107·9 (21)}	{98·9 (30)}	{99·9 (30)}	{100·5 (34)}	{102·3 (87)}
100 *NB/NH*	48·8 (60)	47·2 (53)	49·0 (23)	49·3 (40)	49·4 (32)	50·4 (124)	48·0 (73)
100 O_2/O_1', *R*	90·6 (57)	91 5 (43)	88·1 (22)	91·4 (39)	91·7 (30)	87·8 (120)	—
100 O_2/O_1', *L*					92 4 (29)		—
N∠	—	—	—	{66°·2 (29)}	{67°·0 (23)}	{65°·7 (25)}	—
A∠	—	—	—	{69°·5 (29)}	{69°·6 (23)}	{72°·0 (25)}	—
B∠	—	—	—	{44°·3 (29)}	{43°·4 (23)}	{42°·3 (25)}	—

* All the series in this table are based on Dr Hrdlička's material except the Pecos Pueblo measured by Prof. Hooton. There are more characters available for the last than those given here, but the additional ones (which include the principal calvarial arcs) were not recorded for any other material treated in this paper.

The majority of the skulls for which measurements are given in the 1931 section of Hrdlička's *Catalogue* are artificially deformed, and the two following series are the only ones of undeformed specimens which can be taken from it.

(*d*) *Basket-maker*, 33. These skulls of cave-dwellers in southern Utah are said to form a remarkably uniform collection which cannot be subdivided into types. A few of the Shoshonean specimens came from Utah.

(*e*) *Old Zuñi*, 35. These skulls were collected in Havikuh village in New Mexico. According to Hrdlička's classification, they and the Basket-makers represent the "dolichoid group" of the Pueblo peoples. There are no sufficiently long series of undeformed crania in his *Catalogue* representing the "brachycranic" Pueblo group. This last is represented by the following series recorded by Professor Hooton.

(*f*) *Pecos Pueblo*. The total is divided into four groups representing different archaeological strata, and the whole period represented is probably rather less than one thousand years. The majority of the specimens in each subseries are artificially deformed. Means, standard deviations and coefficients of variation are provided for the "deformed" and "undeformed" crania in each archaeological group, but the numbers for the latter kind are so small that a series long enough for statistical purposes can only be obtained by pooling them. For most of the facial characters the only constants provided are for the total series, as it was assumed that these had not been modified by the calvarial deformation. The means given in our Table IX are thus based on a short series of forty-six skulls in the case of the calvarial measurements and on a longer one of 126 skulls (including the forty-six) in the case of most of the facial characters. This is unfortunate, as there is a possibility that the undeformed people did not represent a random sample from the total population, and comparison of their mean facial measurements with those of the deformed series would have been of interest. As individual measurements are not provided it is not possible to investigate this question. It is shown in section 8 below that the Pecos Pueblo sample is peculiarly heterogeneous when compared with all the others considered, and hence it is not suitable for comparative purposes. Unusual variability is actually exhibited by its calvarial rather than by its facial measurements.

(*g*) *Florida*. In his 1922 publication Hrdlička gives individual measurements of a considerable number of crania of Florida Indians, from mounds and shell-heaps, divided into several short series. A few of these are artificially deformed, and measurements suspected to have have been affected are enclosed in brackets. All measurements not distinguished in this way were included by the author, and by us, in computing means, although those of a few slightly deformed specimens are thus used. He gives pooled means for all the skulls except those of Seminoles, for this group omitting the Seminoles divided into two sub-groups—one being composed of all the specimens with cephalic index above 80 and the other of those with the index below 80—and for the Seminoles separately. The division

on the basis of the cephalic index is a purely arbitrary procedure. The small Seminole series of eleven male skulls was apparently kept separate not because these specimens are clearly distinguished from the others on account of their appearance or measurements, but because the Seminoles are believed to have had a rather different origin from the other natives of Florida. The mean measurements do not lend support to this view. In order to obtain a larger series, the total material was divided into two, one being made up by all the crania from the west coast and the other by the remainder which includes the Seminole specimens. The male means found for these two groups are:

	L	*B*	*H'*	100 *B/L*	100 *H'/L*
West coast	179·9 (78)	145·3 (78)	141·4 (55)	80·8 (78)	79·1 (55)
Others	180·0 (43)	143·4 (43)	141·3 (32)	79·7 (43)	78·2 (32)
	G'H	*J*	*NH*	*NB*	100 *NB/NH*
West coast	74 7 (44)	140·9 (40)	52·9 (47)	25·0 (47)	47·5 (47)
Others	72·7 (21)	139·4 (25)	52·1 (28)	25 3 (26)	48 8 (26)

By supposing that the standard deviations of the series are of the usual order, all the differences between these two sets of means are found to be insignificant. Accordingly they were combined and the pooled means, given in Table IX, were used for comparative purposes. It is shown below that the variability of this pooled series is quite unexceptional. We do not mean to assert that the Indian population of Florida represented was perfectly homogeneous from a racial point of view, but only that from the evidence available there appears to be no justification for partitioning it More abundant material might make it possible to distinguish subsections of the population which could be differentiated. The absolute measurements for which means are given above are the only ones provided for the Florida skulls, with the exception of the total facial height from nasion to menton, and the series is metrically described less adequately than all the others dealt with in this paper.

The reduced coefficients of racial likeness for all possible pairs of these seven series from Western and Southern States, and between them and the Californian and Algonkin series, are given in Table X, and they are discussed in the following section. The Pecos Pueblo series is included here although it is considered to be unsuitable for comparative purposes on account of its exceptionally great variability.

TABLE X

*Reduced coefficients of racial likeness for series of male crania from Western and Southern States and between them and Californian and Algonkin Series**

	$\bar{n}$	Sioux	Arikara	Shoshonean	Basket-maker	Old Zuñi	Pecos Pueblo	Florida
		59·7	49·1	22·3	34·4	30 1	71·7	88·6
Sioux	59·7	—	12·26 ±0 46 (15)	19 92 ±0·76 (15)	90·42 ±0·56 (15)	96 55 ±0·60 (15)	63·31 ±0·36 (15)	98·17 ±0·40 (11)
Arikara	49·1	12·26 ±0·46 (15)	—	24·55 ±0·80 (15)	66 29 ±0 60 (15)	51·86 ±0·64 (15)	38·72 ±0·40 (15)	37·44 ±0·44 (11)
Shoshonean	22·3	19·92 ±0·76 (15)	24·55 ±0·80 (15)	—	36·44 ±0 90 (15)	44·23 ±0 95 (15)	31·72 ±0·70 (15)	92·48 ±0 80 (11)
Basket-maker	34·4	90 42 ±0·56 (15)	66·29 ±0·60 (15)	36·44 ±0 90 (15)	—	4·74 ±0·70 (18)	29·38 ±0·48 (18)	100·65 ±0·58 (11)
Old Zuñi	30·1	96·55 ±0 60 (15)	51 86 ±0·64 (15)	44·23 ±0 95 (15)	4·74 ±0·70 (18)	—	15·68 ±0·53 (18)	68·73 ±0·61 (11)
Pecos Pueblo	71·7	63 31 ±0·36 (15)	38·72 ±0·40 (15)	31·72 ±0·70 (15)	29·38 ±0·48 (18)	15·68 ±0·53 (18)	—	18 67 ±0·35 (11)
Florida	88·6	98·17 ±0·40 (11)	37·44 ±0·44 (11)	92·48 ±0·80 (11)	100·65 ±0·58 (11)	68·73 ±0·61 (11)	18·67 ±0·35 (11)	—
Californian series:								
Northern	48·1	73·17 ±0·53 (11)	41·91 ±0·57 (11)	39·05 ±0·93 (11)	49 04 ±0 64 (14)	29·11 ±0·69 (14)	15·51 ±0·46 (14)	28·70 ±0·46 (11)
Central (Yokuts)	41 6	53·94 ±0·58 (11)	21·75 ±0·62 (11)	55·87 ±0·98 (11)	59·21 ±0·69 (14)	42·64 ±0·73 (14)	16·92 ±0·50 (14)	9·42 ±0·50 (11)
San Francisco Bay and vicinity	82·7	66·01 ±0·35 (15)	38·58 ±0 39 (15)	25·34 ±0·69 (15)	16·34 ±0·46 (18)	15·04 ±0 51 (18)	17·30 ±0·29 (18)	47·02 ±0·29 (11)
Santa Barbara County	43·7	64·81 ±0·49 (15)	50·07 ±0·53 (15)	18·32 ±0·83 (15)	27·21 ±0·63 (15)	19·71 ±0·68 (15)	27 27 ±0·43 (15)	65·26 ±0·48 (11)
Santa Cruz and Santa Rosa	157·8	52·51 ±0·28 (15)	48·74 ±0·32 (15)	15·46 ±0·62 (15)	41·23 ±0·40 (18)	35 70 ±0·44 (18)	42·74 ±0·23 (18)	83·96 ±0·24 (11)
Santa Catalina, etc.	35·4	46·91 ±0·59 (13)	61·67 ±0·63 (13)	24·18 ±0·96 (13)	44·47 ±0·69 (16)	74·17 ±0·73 (16)	84·25 ±0·52 (16)	174·98 ±0·54 (11)
Algonkin series:								
North-Eastern States	91·5	68·33 ±0 34 (15)	42·72 ±0·38 (15)	44·02 ±0·69 (15)	27·92 ±0 48 (15)	46·73 ±0·53 (15)	34·79 ±0·29 (15)	86·50 ±0·31 (11)
East-Central States	58·5	76·53 ±0·42 (15)	33·87 ±0·46 (15)	56·65 ±0·76 (15)	39·28 ±0·56 (15)	51·50 ±0·60 (15)	19·82 ±0·38 (15)	41·21 ±0 39 (11)
Kentucky	27·9	137·43 ±0 65 (15)	82·59 ±0·69 (15)	78·48 ±0·99 (15)	42·43 ±0 79 (15)	27·74 ±0·84 (15)	23·21 ±0·59 (15)	45·80 ±0·67 (11)
Western States	44·1	15·80 ±0·49 (15)	4·37 ±0·53 (15)	17·06 ±0·83 (15)	53·50 ±0·63 (15)	52·51 ±0·67 (15)	26·79 ±0·43 (15)	37·29 ±0·48 (11)

* The $\bar{n}$ given for a particular series in this table is the average number of skulls available for the coefficient of racial likeness characters in the case of the largest number of these characters which can be used. For other numbers of characters the $\bar{n}$'s are different, but the divergences are small in all cases.

6. THE RELATIONSHIPS OF THE NORTH AMERICAN INDIAN SERIES JUDGED FROM THE COEFFICIENTS OF RACIAL LIKENESS AND A COMPARISON OF SINGLE CHARACTERS

All the reduced coefficients of racial likeness having values less than 19 found between the sixteen series—omitting the Pecos Pueblo—are indicated in Fig. 1. This also indicates approximately the localities from which the series were obtained. As has been pointed out in describing the material, some of these areas overlap to a considerable extent in the case of the Shoshonean, Sioux, Arikara and Western Algonkin collections. There are two of the series showing no connections of the order considered, viz. the Californian from Santa Catalina and two neighbouring islands and the Kentucky Algonkin. The former is obviously of a specialised type, as its calvarial length and cephalic and height-length indices are close to the extremes for all races in the world. It is not unusual to find that an island population is of a distinctive type. The Kentucky Algonkin series is chiefly distinguished on account of the small size of nearly all its means of absolute measurements, and it is to be expected that some close connections with it would be found if more material were available for neighbouring populations.

In a general way, the closest resemblances are found between neighbouring peoples, as has been found in the comparison of other similar groups, but there are several exceptions to this. The most striking is found in the case of the four Algonkin series, which are remarkably dissimilar in type, and it must be concluded that the linguistic grouping has little ethnic significance. The close resemblance of the Florida and Central Californian types is also unexpected. It is shown in section 7 below, from comparisons with Asiatic material, that it is safer to neglect some of the higher reduced coefficients shown in Fig. 1. If no account is taken of any greater than 13, then the Shoshonean series also becomes isolated and the Basket-maker and Old Zuñi are detached from the Californian.

What we can assert at present is that there are marked divergences between various Indian tribes of the United States. How many races should be recognized cannot, we feel, be stated with precision. It may not be amiss to point out, however, that Professor von Eickstedt's classification (1934, pp. 678–88) is fairly well in accord with our findings. His "margide Gruppe" is represented by the Californian and Florida series. It is particularly interesting to find a bond between the two. His "sylvide Gruppe", however, has to be broken up into at least two: the prairie Indians, represented by the Sioux and the Arikara, and the Indians of the north-eastern forests, the Eastern Algonkin. The Kentucky group, which is found to be isolated, indicates that there may be further races of which we have insufficient knowledge at present. The "zentralide Gruppe" may be represented by the Old Zuñi and Basket-makers, but these are dolicho- or mesaticephalic (73·0 and 75·9, respectively), certainly not brachycephalic, as

von Eickstedt describes his group. It may be pointed out in passing that these two types resemble quite closely that of the Peruvian skulls described by MacCurdy (1923).

In making comparisons between the mean measurements for a number of cranial series, the relative extents to which different characters differentiate them can be estimated conveniently by comparing the percentages of significant differences found. The question whether a particular difference is significant or not can be judged from the α obtained for it in computing the coefficient of racial likeness. An α is approximately the square of a quantity which is the difference of two means divided by its standard error, and it may be arbitrarily supposed to indicate differentiation if it is greater than 10. The percentages of α's found greater than 10 are given below for all the comparisons between the sixteen American Indian series—omitting the Pecos Pueblo—and for all the comparisons between 12 Oriental series:*

	100 H'/L	100 B/H'	100 B/L	H'	B	L	NH	O_2	NB
16 American series	74 2	68·3	65 8	64·2	60·8	54·2	51·7	48·7	48·3
12 Oriental series	35·9	40·6	57·6	35·9	28·8	51·5	49·1	21·8	48·5

	J	C	$G'H$	O_1' (or O_1)	$N\angle$	100 O_2/O_1'	LB	100 NB/NH	$A\angle$
16 American series	47·5	44·9	39·2	37 9	33·3	27·3	14·3	10·0	0
12 Oriental series	13·6	0	48·5	34·0	14·3	22·6	24·2	55·4	10·7

These two sets of frequencies arrange the characters in rather dissimilar orders. For the American series the highest percentages are shown by the three major calvarial indices and the three diameters from which these are obtained: for the Oriental series these characters also give percentages among the highest, but they are equalled or exceeded by those for the three nasal measurements and the upper facial height. For all of the 18 characters except NB, $G'H$, $100NB/NH$, LB and $A\angle$ the American percentage is greater than the corresponding Oriental value, and in several cases markedly greater. There is, in fact, marked diversity among the Indian types of the United States compared with that normally found for comparable groups in other parts of the world.

* These last percentages have been given by Woo and Morant (1932, pp. 130–1).

The arrangements provided by the means for single characters, or pairs of characters, are far less suggestive than that given by the coefficients of racial likeness, and detailed consideration of them would not be profitable.

7. COMPARISONS OF NORTH AMERICAN INDIAN WITH ASIATIC AND ESKIMO CRANIAL SERIES

In preceding sections of this paper the coefficients found between all possible pairs of sixteen male series of crania representing Indian populations of the United States have been given. Interpretation of these generalised measures of the resemblances of the types will obviously be aided if the results of comparisons of the same kind between the Indian and other groups of series are also known. Such intergroup comparisons were made with Asiatic and Eskimo material.

The coefficients have been given for all pairs of 26 male Asiatic series (Woo & Morant, 1932) and for all pairs of seven Eskimo series (Morant, 1937). In the paper on the latter comparisons were also made between them and the Asiatic series, though actually only one coefficient of this kind is given. Computation in full of the remaining 181 ($182 = 26 \times 7$) was considered unnecessary because a test applied showed that all these reduced coefficients were extremely likely to be greater than 19, and no account was taken of values greater than this in obtaining the classification of the Asiatic series. The test in question depends on the fact that for these groups the calvarial length, breadth and height and the three indices derived from these measurements gave percentages of significant differences (indicated by α's greater than 10) larger than, or almost as large as, the percentages given by any other of the 31 characters used. The values of the coefficients were evidently determined largely by these six measurements, and it has been shown in section 6 above that the same is true for the North American Indian series. For the two groups of series the maximum differences between the means found in the case of comparisons which give reduced coefficients of racial likeness less than 19 are:

	L	B	H'	$100\,B/L$	$100\,H'/L$	$100\,B/H'$
26 Asiatic series	6·7	6·1	6·3	5·4	3·4	6·5
16 North American Indian series	7·4	7 2	6·5	3·6	3·5	5 0

Considering the Asiatic series alone, if any one of the 26 available could be compared with a new Asiatic series, and if one or more of the differences of the means for the six characters were found to exceed the limit given above, then it is unlikely that the reduced coefficient found would be less than 19. In the same circumstances, it is still less likely that one of the Asiatic and a non-Asiatic

series would give a reduced coefficient less than 19. These considerations make it possible to select, by merely finding the differences of a few means, those pairs of series in new comparisons which will almost certainly provide reduced coefficients greater than the limit (19) arbitrarily chosen. The ranges of the differences actually used for this purpose were those above for the Asiatic series with the addition of 0·1 to each, viz. L 6·8 mm., B 6·2 mm., H' 6·4 mm., $100B/L$ 5·5, $100H'/L$ 3·5 and $100B/H'$ 6·6. After the pairs of series which will probably give coefficients which will indicate greater dissimilarity than any to be used in the classification have been selected in this way, we are left with a number of pairs which may or may not give reduced values less than 19. It is not necessary to calculate all these in full, since it can often be seen from the α's for a few characters only that a value greater than 19 will be obtained, so that there is no need to complete the computation.

The twenty-six Asiatic give 416 comparisons in pairs with the sixteen North American Indian series. The test described shows that 318 of these coefficients are almost certainly greater than 19. Of the remaining ninety-eight, seventy-six were also found to be greater than the limit and it was not necessary to calculate them in full in order to be sure of this. The twenty-two reduced coefficients less than 19 are given in Table XI. It should be noted that connexions of the order considered are only found between seven Oriental types and the Chukchi, on the one hand, and twelve of the sixteen American types, on the other. Most of the southern Oriental series, all the Northern Mongolian (Siberian) and all the Indian series are excluded. The fact that the closest resemblances are between eastern and north-eastern Asiatic and the American populations is in accordance with expectation, but a moment's consideration shows that little significance can be attached to the measures of resemblance which lead to this conclusion. The American series are thereby connected with the Oriental in what appears to be a haphazard way. For example, the Kentucky Algonkin series is linked to the Japanese (reduced coefficient = 17·0), while the lowest coefficient found between it and the fifteen other American series is 27·7: the North-Eastern and East-Central Algonkin series were found to be connected only with one another when comparisons were confined to American material, but the former shows a connection with the Aino and the latter with the Japanese, Aino and Chinese Prehistoric series. Results such as these can only be considered so unreasonable that the assumption that the method used is capable of presenting the situation in such a way that it will be possible to unravel the skein of interrelationships seems to be discredited.

There is the possibility, however, that the defect is due not to the method in itself but to the way in which it is being used. The fact that different *high* orders of reduced coefficients are not capable of indicating different degrees of distant relationship can easily be demonstrated. Hence it was concluded that only values below a certain limit should be considered. The limit chosen in the case of

TABLE XI

Reduced coefficients of racial likeness less than 19 between North American Indian and Asiatic series of male skulls

	$\bar{n}$*	Dayak	Middle Java	Tibetan A	Fukien Chinese
		48·2	64·4	35·9	36·0
Northern California	48·1	18·52 ± 0·53 (14)	16·83 ± 0·46 (14)	16·75 ± 0·62 (14)	15·16 ± 0·62 (14)
Santa Barbara, California	43·7	—	—	16·24 ± 0·62 (15)	—
San Francisco Bay	82·7	—	—	—	14·51 ± 0·45 (18)
Central California	41·6	—	—	—	15·67 ± 0·66 (14)
Florida	88·6	—	—	—	18·01 ± 0·56 (11)

	$\bar{n}$*	Japanese	Aino	Chukchi	Chinese Prehistoric
		118·7	80·7	34·0	39·1
Northern California	48·1	17·47 ± 0·39 (12)	—	—	—
San Francisco Bay	82·7	12·18 ± 0·24 (16)	10·82 ± 0·28 (16)	—	—
Central California	41·6	—	—	11·86 ± 0·73 (12)	—
Kentucky, Algonkin	27·9	17·00 ± 0·54 (15)	—	—	—
North-Eastern Algonkin	92·5	—	16·43 ± 0·31 (13)	—	—
Western Algonkin	44·8	—	15·77 ± 0·46 (13)	6·83 ± 0·71 (12)	—
Shoshonean	22·6	—	—	18·50 ± 1·01 (12)	—
Arikara	50·6	—	—	2·25 ± 0·68 (12)	—
East-Central Algonkin	58·9	13·26 ± 0·31 (15)	15·96 ± 0·38 (13)	—	17·73 ± 0·57 (12)
Old Zuñi	31·5	—	—	—	18·09 ± 0·76 (13)

* Where more than one coefficient is given for a particular series, the $\bar{n}$ for it in this table is the mean number of skulls available for the coefficient of racial likeness characters in the case of the coefficient based on the largest number of characters.

comparisons of Asiatic series with one another was 19, because the arrangement provided by all the values less than 19 appeared to be a reasonable and suggestive one for them. The same may be considered true, as far as can be seen, for the North American Indian series considered by themselves (see Fig. 1), but this is not so when the cross connections between the Asiatic and American series are considered. But it is still possible that the limit chosen is really too high and that more reasonable results would be obtained if it were lowered.

Before considering this question the results of comparisons between the American Indian and Eskimo series may be given. There are sixteen in the former and seven in the latter group and a comparison of the six calvarial measurements suggested that seventy-eight of the 112 comparisons would give reduced

coefficients of racial likeness greater than 19. It was found that thirty-one of the remaining thirty-four comparisons also give values above the same limit, leaving the following three reduced coefficients: Western Eskimo (220·0) and Arikara (49·1)—7·07 ± 0·31 (15); Western Eskimo (220·0) and Western Algonkin (44·1)—15·91 ± 0·33 (15); Point Hope Eskimo (125·1) and East-Central Algonkin (58·5)—17·32 ± 0·31 (15).

The data in Table XI show that any attempt to take into account all reduced coefficients less than 19 is likely to be unprofitable when considering the classification of the three groups of races considered. All the connexions between the series which remain when the limit is reduced to 13 are shown in Fig. 2, and this new limit has again been chosen arbitrarily merely because it appears to lead to the most suggestive arrangement. The position as far as the United States Indian series considered alone are concerned is little changed, except that the Shoshonean series has become isolated, and that the Basket-maker and Old Zuñi lose their connexions with the Californian series. The arrangement of the Eastern Asiatic series is less changed and the continuous system which they form remains intact. There are only two connexions between the two groups, viz. those linking the San Francisco Bay series to the Aino and Japanese. There are closer relations between the American Indian types and those of the Western Eskimo and Chukchi populations, and these are somewhat unexpected in view of the fact that no data from Canada are available.

The evidence suggests forcibly that in attempting to estimate relationships by these methods it will be safest to ignore all reduced coefficients of racial likeness greater than 13, as inconsistent results are likely to be obtained if significance is attached to differences indicating more distant degrees of resemblance. This restriction actually makes it necessary to discard certain suggestions which appeared to be of considerable interest. For example, if significance is attached to any reduced coefficient less than 19 then the Chukchi is found to have only one connexion with the Asiatic series* (viz. with the Chinese Prehistoric), and only one with the Eskimo (viz. with the Western Eskimo). A link is thus found between the two groups precisely where it would have been expected. But there can be no justification for accepting this result and at the same time refusing to interpret the evidence of the majority of the coefficients in Table XI in the same way.

It should be decided, then, that the most suggestive classification is likely to be reached if no account is taken of any reduced coefficients greater than 13, but this limit may still be too high. If it is reduced to 10, say, the arrangement shown in Fig. 2 is broken up, as it were, into a few constellations of series having no connexions with one another, and in the case of the American Indian material of a number of isolated series as well. It may be anticipated that these last would become linked to one another to form a constellation if more series from the area were available, but no demonstration of this can be given at present. If the view

* Excluding the short Tibetan *B* series which gives a reduced coefficient of 14·5 with the Chukchi.

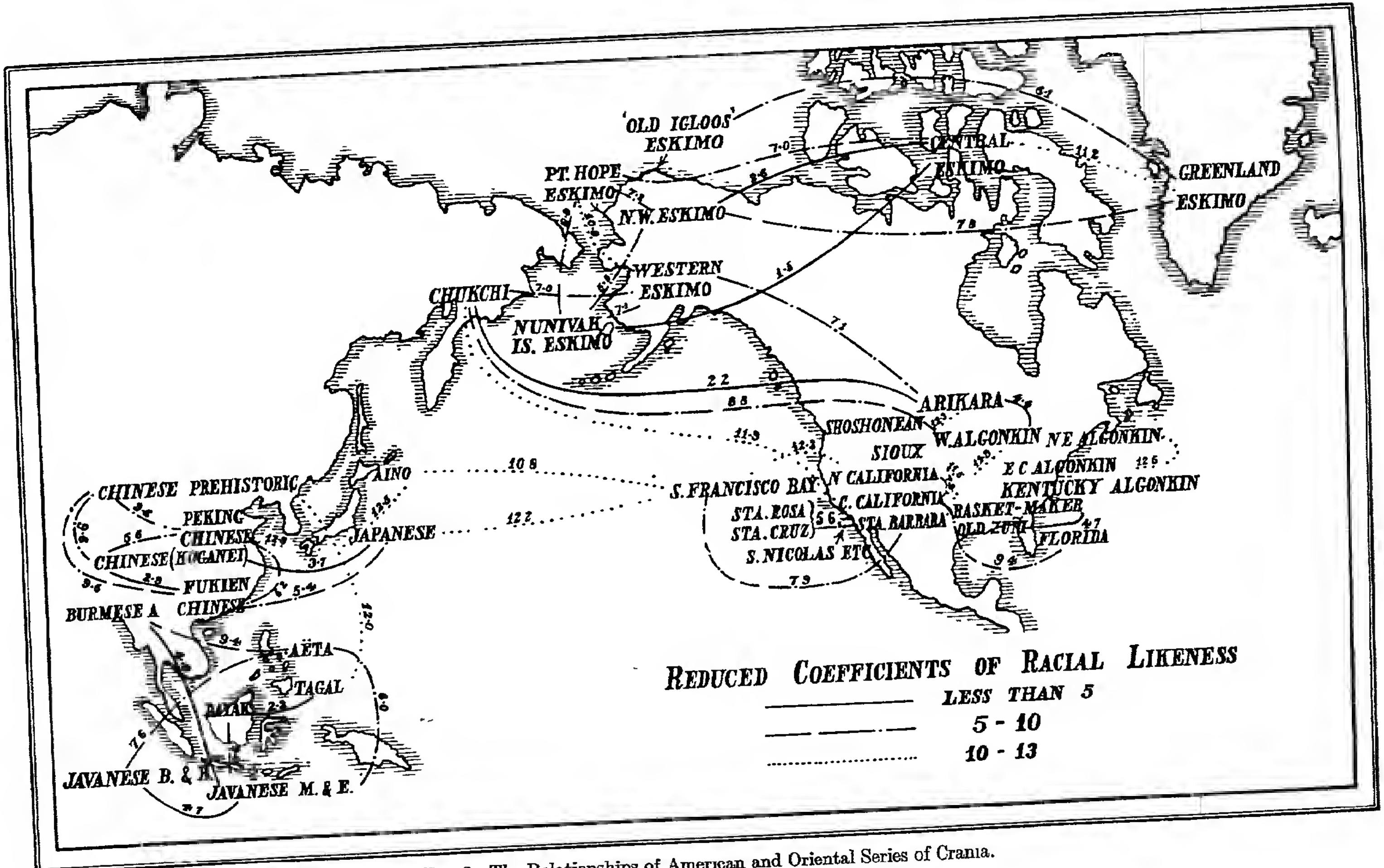


Fig. 2. The Relationships of American and Oriental Series of Crania.

that a reliable classification can only be based on the evidence of the close resemblances of types be correct, then it is clear that a considerable number of series representing a particular group of races must be available before it becomes possible either to estimate their interrelationships, or to determine the links between them and other groups. In the present case it is safest to conclude that the cranial material available makes it possible to distinguish a few groups of closely allied peoples among those of Eastern Asia and North America, but that the connexions between the groups, and the affinities of a number of types which do not fall within them, must remain undecided until new material clarifies the position.

The links found between a Californian type and the Aino and Japanese are suggestive, but little importance can be attached to them at present. The results of the cranial comparisons appear to be in favour of the hypothesis which postulates an immigration into the American continent via the Straits of Bering. The Chukchi may then be considered as a tribe left in Asia during this migration, and the resemblance between the Arikara and the Western Eskimos may be taken as an indication of a former contact between the two races. The links between the Californian and the Oriental types may be an indication of the same route of migration rather than a sign of direct trans-Pacific traffic. Japanese and American Indians are too far removed as regards the colour of their skin and other integumentary characters to make this link appear plausible. It must be remembered that neither the Japanese nor the Californian series of skulls used in these comparisons are adequately described by the measurements given for them, and better material might lead to rather different conclusions.

It may be noted that there are no characters for which means are available which distinguish all the Asiatic from all the North American Indian types, though many of the most significant differences between them are due to the fact that the latter tend to have the broader and higher facial skeletons. The length, breadth and height of the brain-box and the three indices derived from these measurements are remarkably similar for some of the pairs of series, but in the cross comparisons only one case was found—viz. the Northern Californian compared with the Fukien Chinese series—for which all the differences of the means for the six calvarial characters are insignificant. The same is found for the Western Eskimo, on the one hand, compared with the Arikara and Western Algonkin, but all the Eskimo types have decidedly lower nasal indices than all the North American Indian.

8. The variabilities of series of North American Indian skulls

In the foregoing sections of this paper comparisons are made between the means of 17 male series of crania which were finally selected for the purpose, and all these relate to Indian populations of the United States. The means for these are given in Tables I, VI and IX. The Kentucky (Table VI) and Shoshonean (Table IX) series were considered to be too short to give estimates of any value

of the variabilities of the populations they represent. The standard deviations for the remaining 15 series are provided in Table XII, omitting a few values which can only be given for fewer than thirty specimens.

In comparing these constants two groups of the series will first be considered separately. The first consists of six from California. Taking each character in turn, the differences between all possible pairs of the standard deviations were estimated in terms of their probable errors, and these ratios will now be supposed to indicate significant deviations if they are greater than 3·5. For 9 of the 14 characters concerned the constants for the Californian series show no differences of this order, for L there are found to be 2 significant differences, for NH 2, for B 4, for $G'H$ 4 and for 100 B/L 5. Of the total seventeen ratios greater than 3·5, the largest is 5·5 and there are only four greater than 5·0. It should be remembered that in a set of ratios of the kind considered some values greater than the limit chosen must be expected owing to chance. In comparing different pairs of the Californian series, the numbers of characters which can be used range from 8 to 11. There are four pairs of series showing no significant differences, six showing 1 only, four showing 2 only and one showing 3 significant differences only. It is clear that there is no evidence to show that the six Californian populations represented differed substantially in variability, and, in view of the danger that the small samples available may not have been drawn entirely at random, it appears safest to conclude that these populations all exhibit the same degree of variation.

The second group of series referred to is made up by the following eight in Table XII, viz. all the "Other U.S.A. series" except the Pecos Pueblo. A similar treatment leads to precisely the same conclusion in this case. In a total of 325 comparisons, the difference exceeds 3·5 times its probable error in thirty-two instances, there are three ratios greater than 5 and the largest is 7·2. Comparing the series in pairs—the numbers of characters which can be used range from 9 to 13—there are found to be eight pairs showing no significant differences, eleven showing 1 only, six showing 2 only and three showing 3 significant differences only. A close approach to equality in variation is again indicated.

The Pecos Pueblo series was not included in the second group because its standard deviations are obviously peculiar. For 7 of the 13 characters they are greater than all the corresponding values for the other fourteen series. Little importance can be attached to the fact that the Pecos Pueblo standard deviation is extreme in the case of the capacity (C) and orbital index (100 O_2/O_1'), but this is not so for the remaining 5 characters for which the situation (for fourteen comparisons) is:

L, Pecos Pueblo σ significantly greater than 12 others and highest ratio 6·6;

B, Pecos Pueblo σ significantly greater than 3 others and highest ratio 5·5;

H', Pecos Pueblo σ significantly greater than 4 others and highest ratio 5·2;

100 B/L, Pecos Pueblo σ significantly greater than 6 others and highest ratio 6·1;

TABLE XII

Standard deviations of North American Indian series of male skulls

	Californian Series					
	Northern California	Central California	San Francisco Bay and vicinity	Santa Barbara County	Santa Cruz and Santa Rosa Islands	Santa Catalina, San Clemente and San Nicolas Islands
C	—	—	—	—	110·2 ±4·6 (128)	—
L	6·85±0·44 (54)	5·87±0·42 (45)	5·72±0·23 (146)	4·59±0·32 (48)	5·45±0·19 (195)	4·68±0·35 (40)
B	5·49±0·36 (54)	4·68±0·34 (44)	4·75±0·19 (143)	3·37±0·24 (45)	4·58±0·16 (192)	5·45±0·41 (41)
H'	5·46±0·36 (51)	5·82±0·43 (42)	4·55±0·21 (103)	4·65±0·33 (45)	4·60±0·16 (188)	4·48±0·36 (35)
LB	3·48±0·24 (49)	3·39±0·25 (41)	3·76±0·21 (75)	—	3·81±0·18 (102)	4·06±0·33 (34)
GL	4·87±0·36 (42)	4·00±0·31 (39)	4·96±0·32 (54)	—	4·38±0·21 (98)	4·67±0·40 (31)
$G'H$	3·83±0·28 (44)	6·04±0·43 (44)	4·34±0·21 (93)	3·37±0·23 (47)	3·57±0·12 (191)	4·11±0·33 (36)
J	5·89±0·44 (40)	6·22±0·52 (32)	6·45±0·36 (73)	—	5·19±0·19 (172)	4·97±0·38 (38)
NH	3·19±0·22 (48)	4·10±0·29 (45)	3·02±0·14 (101)	2·64±0·18 (49)	2·77±0·09 (199)	3·03±0·23 (40)
NB	1·72±0·12 (48)	1·77±0·13 (44)	1·80±0·09 (101)	1·88±0·13 (51)	1·50±0·05 (199)	1·98±0·15 (40)
O_2	—	—	—	1·95±0·14 (45)	1·70±0·07 (141)	—
O_1'	—	—	—	1·36±0·10 (44)	1·37±0·07 (86)	—
$100\ B/L$	4·21±0·28 (53)	2·42±0·18 (42)	3·24±0·13 (139)	2·43±0·17 (44)	2·90±0·10 (191)	3·21±0·24 (40)
$100\ NB/NH$	5·10±0·35 (47)	5·01±0·36 (44)	4·16±0·20 (98)	3·86±0·27 (48)	4·02±0·14 (199)	4·90±0·37 (40)
$100\ O_2/O_1'$	—	—	—	4·51±0·32 (44)	4·01±0·21 (86)	—

TABLE XII (*continued*)

	Other U.S.A. Series					
	North-Eastern Algonkin	East-Central Algonkin	Western Algonkin	Sioux	Arikara	Florida
C	—	75·1 ±6 0 (36)	93·1 ±7·0 (40)	95 1 ±6 0 (57)	98·4 ±7·2 (43)	—
L	5·19±0·23 (120)	5·61±0 28 (94)	4·95±0·34 (49)	5·60±0·34 (63)	5·20±0 35 (51)	5·26±0 23 (121)
B	4·23±0·18 (120)	4·72±0·23 (93)	5·13±0·35 (49)	4·70±0·28 (63)	4 27±0·29 (51)	5·63±0·24 (121)
H'	5·00±0·23 (110)	4·48±0·27 (64)	5·65±0·39 (47)	3·53±0·22 (61)	4·95±0·33 (50)	4·39±0·22 (87)
LB	—	—	—	—	—	—
GL	—	—	—	—	—	—
$G'H$	3·44±0·20 (70)	3·58±0·28 (36)	3·67±0·27 (41)	3·97±0·25 (56)	3·43±0·23 (51)	4·19±0·25 (65)
J	5·98±0·33 (77)	5·91±0·51 (31)	4·87±0·39 (35)	4·38±0·27 (59)	4·49±0·30 (52)	6·11±0·36 (65)
NH	2·84±0 14 (90)	2·46±0·16 (51)	2·60±0·18 (47)	2·70±0·16 (61)	2·11±0·14 (53)	2·85±0·16 (75)
NB	2·22±0·11 (89)	2·07±0·14 (52)	1·51±0 11 (46)	2·18±0·13 (60)	1·77±0·12 (53)	1·81±0 10 (73)
O_2	1·59±0·08 (87)	1·58±0·11 (51)	1·54±0·12 (39)	1·75±0·11 (57)	1·98±0·14 (43)	—
O_1'	1·38±0·07 (85)	1·45±0·10 (51)	1·36±0·10 (39)	1·56±0·10 (57)	1·13±0·08 (43)	—
100 B/L	2·95±0·13 (120)	3 07±0·15 (93)	3·31±0·23 (49)	3·03±0 18 (63)	3·20±0·21 (51)	2·93±0·13 (121)
100 NB/NH	4·23±0·21 (89)	4·59±0·31 (50)	4·49±0·32 (46)	4·01±0·25 (60)	3·63±0 24 (53)	3·69±0·21 (73)
100 O_2/O_1'	4·03±0·21 (85)	3·20±0 21 (51)	4·17±0·32 (39)	3·98±0·25 (57)	4·53±0·33 (43)	—

TABLE XII (*continued*)*

	Other U.S.A. Series			Averages for Californian Series	Averages for other U.S.A. Series†	Averages for all U.S.A. Series†	Egyptian E‡
	Basket-maker	Old Zuñi	Pecos Pueblo				
C	76·6 ±6·6 (31)	—	118·1 ±10·1 (31)	110·2 :1 (128)	92·2 :5 (207)	99·5 :6 (335)	113·5 ±2·0 (753)
L	3·84±0 32 (33)	5·42±0·44 (34)	8·15± 0·57 (46)	5·59:6 (528)	5 25:8 (565)	5·42:14 (1093)	5·72±0 09 (895)
B	4·77±0 40 (33)	5·23±0·42 (35)	6 14± 0·44 (45)	4·72:6 (519)	4·87:8 (565)	4·80:14 (1084)	4·76±0·08 (896)
H'	3·58±0 31 (30)	3·49±0·30 (30)	6·49± 0·53 (34)	4·81:6 (464)	4·56:8 (479)	4·68.14 (943)	5 03±0·08 (884)
LB	3·38±0·28 (33)	—	—	3·72:5 (301)	3 38.1 (33)	3 69: 6 (334)	3·97±0·06 (896)
GL	—	—	—	4·57:5 (264)	—	4·57: 5 (264)	4·85±0·08 (832)
$G'H$	4·29±0·36 (33)	3·43±0·29 (32)	3·95± 0·18 (112)	4 08:6 (455)	3·77:8 (384)	3·94:14 (839)	4·15±0·07 (845)
J	4·44±0·35 (36)	3·95±0·34 (31)	6·17± 0 29 (102)	5·63:5 (355)	5·21:8 (386)	5·41:13 (741)	4·57±0 08 (785)
NH	2·50±0 19 (41)	2·28±0 19 (32)	2·74± 0 12 (125)	3·02:6 (482)	2·61:8 (450)	2 83:14 (932)	2·92±0 05 (898)
NB	1·20±0·09 (40)	1·50±0·13 (32)	1·57± 0·07 (126)	1·70.6 (483)	1·89:8 (445)	1·79:14 (928)	1·77±0·03 (893)
O_2	1·35±0·10 (40)	1·29±0·11 (30)	1·61± 0 07 (119)	1·76.2 (186)	1·62:7 (347)	1·67: 9 (533)	1·88±0 03 (888)
O_1'	1·33±0·10 (39)	1 47±0·13 (30)	1·82± 0 08 (119)	1·37:2 (130)	1·39:7 (344)	1·39: 9 (474)	1·65±0·03 (880)
100 B/L	3·28±0 27 (33)	4·08±0·33 (34)	4·81± 0·35 (43)	3·11:6 (509)	3·13:8 (564)	3·12:14 (1073)	2·68±0·06 (884)
100 NB/NH	2·95±0·22 (40)	3·13±0·26 (32)	4·27± 0·18 (124)	4·33:6 (476)	3·95:8 (443)	4·15:14 (919)	3·82±0·06 (881)
100 O_2/O_1'	3·88±0 30 (39)	4·26±0·37 (30)	4·57± 0 20 (120)	4·19·2 (130)	4·00:7 (344)	4·05: 9 (474)	4·95±0 08 (876)

* The average standard deviations given in the three columns preceding the last one were computed by weighting the squares of the standard deviations of the component series, as described in the text. In these cases the number following an average σ, e.g. :6, is the number of series on which it is based, and the number in brackets gives the total skulls involved.

† Excluding the Pecos Pueblo series.

‡ Given by Davin and Karl Pearson. In the case of this series the standard deviations are for the vertical height from the basion (H) instead of for H', for O_2L, for the maximum breadth of the left orbit from the medial margin (O_1L) instead of for O_1' and for 100 O_2/O_1, L instead of for 100 O_2/O_1'. These pairs of measurements give closely similar standard deviations when they are available for the same series.

O'_1 (9 comparisons), Pecos Pueblo σ significantly greater than 6 others and highest ratio 6·1.

The Pecos Pueblo series is evidently appreciably more variable than any of the others, and as it differs from them in this respect it must be supposed unsuitable for purposes of racial comparison. Its peculiarity may be due either to the fact that the measurements selected because they were believed to have been unaffected by artificial deformation were not uninfluenced by this disturbing factor, or to the fact that the population represented was racially more heterogeneous than all the others.

Comparisons were not made between the variabilities of pairs of the series of which the first belongs to the Californian and the second to the other group of series distinguished, as it is of more interest to compare the two groups in another way. The average standard deviations for the six Californian series given in the fourth column from the end of Table XII were obtained by weighting the squares of the constants for the single series with the numbers of skulls on which they are based. The following column gives averages computed in the same way for the eight other Indian series, excluding the Pecos Pueblo. It is clear that these two sets of average values show a much closer approach to equality than is shown by the standard deviations for almost all pairs of the component series. Probable errors for the average constants have not been computed, but it is probable that most of the differences between the two sets for corresponding characters are quite insignificant. For ten characters the Californian values are in excess and for four in defect of the others, but the absolute differences between the constants are all small, and these relations can only be taken to indicate that the Californian populations show a slight tendency to be more variable than other Indian populations of the United States.

The penultimate column of the table gives the average standard deviations, computed in the way described, for all 14 of the Indian series, still excluding the Pecos Pueblo. These may be compared with the values in the last column for Egyptian skulls obtained from a single cemetery at Gizeh used from the 26th to 30th dynasties.* Probable errors for these last are given, and, in view of the total numbers of skulls on which they are based, it will be safe to assume that for corresponding characters the American averages will have probable errors either of the same order or rather greater than the Egyptian. On this assumption, there seems to be no reason to suspect that the differences between the two sets of standard deviations are clearly significant except in the case of the capacity and three orbital measurements (for which the Egyptian values are the greater) and the bizygomatic breadth (J) and cephalic and nasal indices (for which the American values are the greater). But even in these cases the absolute differences between the corresponding constants are small, and the use of the Egyptian values in computing coefficients of racial likeness between American Indian series seems to be sufficiently justified.

* The Egyptian standard deviations are taken from Pearson & Davin (1924).

9. Conclusions

This paper presents a preliminary classification of the Indian races of the United States derived from the mean measurements of groups of undeformed male adult crania. The data provided by Gifford and Hrdlička were found to be the only ones suitable for the purpose. The total 1167 skulls were divided into sixteen series—three being made up by fewer than forty specimens each—for which means are given. Judging from the standard deviations (Table XII) the sixteen selected series indicate a remarkably close approach to equality in intra-racial variability, and only one other (the Pecos Pueblo) had to be rejected because it appears to represent a decidedly more heterogeneous population. The average standard deviations for the sixteen series are found to be remarkably close to those of a long series of late dynastic Egyptian crania, and this order of variability is rather less than that found for modern series of crania from Western Europe.

Comparisons between the types of the series were made by applying the method of the coefficient of racial likeness, the classification suggested being derived solely from the lowest orders of reduced coefficients. When possible these constants are based on thirty-one cranial characters, but for the American material they can only be computed for numbers between 11 and 18, since several of the customary measurements are not available. This limitation is unfortunate, but there is no reason to believe that the *orders* of the reduced coefficients obtained are different from those which would be given if all thirty-one characters could be used.

All the values less than 19 found are indicated in Fig. 1. Comparisons were also made between the sixteen North American Indian series, on the one hand, and Oriental and Eskimo series on the other, and all the reduced coefficients less than 13 within and between these three groups are shown in Fig. 2. There are several other connexions between the United States and Oriental series provided by values between 13 and 19, but it is suggested that no significance should be attached to these, and hence that no account should be taken of reduced coefficients greater than 13 in classifying the American series. Owing to the complexity of the problem, it was to be anticipated that the way in which a generalized criterion of resemblance, such as the coefficient of racial likeness, can best be used to furnish a classification of racial types must be determined empirically. The contention that the most suggestive results are obtained by considering the evidence of close resemblances only is fully sustained by the present investigation, but the limiting order of resemblance which can best be used may have to be modified again in the light of more abundant material. If the evidence of all reduced coefficients less than 13 is taken into account, then the only connexion between the North American and Oriental types are the links between a Californian series and the Aino and Japanese. If it should be found necessary to reduce the limit again—to 10, say—then for the existing material there will be

no connexions between the two groups, though it is probable that some would be provided by populations unrepresented at present. The fact that the Chukchi is closely allied to some North American types but not to any of the available Asiatic types is unequivocal, and there are close bonds between the Western Eskimo and the United States types. A surprising diversity is found among the Indian populations of the country, and this is equally apparent whether the coefficients of racial likeness are considered, or the mean measurements are compared in any more direct way. On this account, it will be necessary to have considerably more material than that available at present in order to reveal their interrelationships in a completely satisfactory way. Comparison of the results obtained already with those which might be derived from more adequate metrical descriptions of the same material is also required.

APPENDIX

NEW SERIES OF AMERICAN INDIAN CRANIA

Shortly after the paper above had been written, measurements were published of new series of American Indian crania excavated from mounds in Fulton County, Illinois (Cole & Deuel, 1937). The report on them is said to be an interim one only, but the measurements provided are more detailed than those for nearly all the United States Indian series described previously. The artefacts found with the skeletons made it possible to distinguish six cultural divisions extending from some pre-Columbian date to the seventeenth or eighteenth century, though no objects suggesting contact with Europeans were discovered.

Our means computed from the individual measurements of male undeformed skulls are given in Table XIII for the following groups:

(i) Mounds 14 and 34 (table facing p. 264)—late in date. It is said that the skulls from these two mounds are "very closely related, permitting the pooling of the craniometric data".

(ii) Mounds 85 and 86 (table facing p. 264)—late in date and following or contemporaneous with (i). It is said that these skulls do not differ markedly from those in the first group.

(iii) All other mounds, viz. 7, 10, 11, 12, 13, 14, 15, 77 and 188 (tables in text) —earlier in date. A number of types are distinguished among these skulls but the total is very small.

Mean measurements for these three groups are given in Table XIII. In the case of the majority of the characters considered there, it is clear that all the differences are quite insignificant, though even if this were so for all of them it would not provide good evidence of identity of type owing to the small sizes of the series. Differences which are probably significant are only found for the basio-bregmatic height (H') and the two indices involving this chord. But little stress can be laid on this fact, as so few individuals are represented that the samples are particularly unlikely to be truly random ones representing large populations or

TABLE XIII

*Mean measurements of series of male crania from Fulton County, Illinois, and standard deviations for the total series**

	L	B	H'	LB	B'	J	$G'H$
Mounds 14 and 34	180·1 (27)	140·0 (27)	145·6 (24)	105·5 (24)	94·6 (27)	140·4 (22)	75 0 (26)
Mounds 85 and 86	182·5 (13)	137·3 (13)	140·7 (12)	105·3 (12)	92·8 (13)	136·5 (13)	74·0 (13)
Other mounds	183·2 (18)	140·1 (17)	138 5 (8)	102·3 (8)	94·8 (18)	140·0 (12)	73·0 (12)
Total series	181·6 (58)	139·4 (57)	143·0 (44)	104·8 (44)	94·3 (58)	139·2 (47)	74·5 (51)
σ's for total series	6·77±0·42	4·40±0 28	5·24±0·38	4·98±0·36	—	5 86±0 41	3·57±0·24

	NH	NB	O_1L	O_2L	G_1'	G_2	100 B/L
Mounds 14 and 34	53·5 (27)	27·0 (26)	43·5 (23)	34·4 (26)	47 9 (26)	40·4 (24)	77·8 (27)
Mounds 85 and 86	53·2 (13)	26·0 (13)	41·8 (13)	34·8 (13)	47·9 (13)	40·1 (12)	75·4 (13)
Other mounds	53·3 (12)	26·0 (12)	43·0 (11)	34·9 (12)	47·4 (8)	39·7 (8)	76·8 (17)
Total series	53·4 (52)	26·5 (51)	42·9 (47)	34·6 (51)	47·8 (47)	40·2 (44)	76·9 (57)
σ's for total series	3 10±0·21	1·87±0·12	1·92±0·13	1·95±0·13	—	—	3·83±0·24

	100 H'/L	100 B/H'	100 NB/NH	100 $O_2/O_1, L$	100 G_2/G_1'	*Prosth.* $P\angle$
Mounds 14 and 34	81·2 (24)	{96 2 (24)}	50 4 (26)	79·5 (23)	84·5 (24)	83°·2 (27)
Mounds 85 and 86	77·3 (12)	{97·6 (12)}	49·1 (13)	83 2 (13)	84·8 (12)	84° 5 (13)
Other mounds	76·5 (8)	{101·2 (8)}	49·0 (12)	80·4 (13)	83·7 (7)	84°·0 (12)
Total series	79·3 (44)	{97·5 (44)}	49·7 (51)	80·7 (49)	84·5 (43)	83°·7 (52)
σ's for total series	4·55±0·30	—	—	4·39±0·30	—	—

* The measurements of the Fulton skulls were determined by using Martin's definitions. The symbols for them used here and listed in the footnote to p. 95 above may be taken to indicate exact correspondence with the definitions followed in determining the measurements of the other series used in this paper. Those in this table not available for any of the other series are: B' = minimum frontal breadth (Martin's No. 9) O_1L = breadth of orbit from maxillo-frontale (51), G_1' = length of palate from staphylion to orale (62), G_2 = breadth of palate between the mid-points of the inner alveolar walls of the second molars (63) and *Prosth.* $P\angle$ = angle between chord joining nasion to prosthion and the Frankfort horizontal plane (72).

a single large population. Owing to the limitations of the evidence, it appeared best to pool the three series in the hope that it might represent a single racial group, and accordingly the means and standard deviations for the total series given in the table were computed. Comparisons of its variabilities, however, show that this sample cannot be supposed racially homogeneous. These can be made with the average standard deviations for 14 cranial series given in Table XII in the case of 13 characters, supposing that the variabilities of the orbital breadth and indices determined in the different ways are comparable, since they give almost identical standard deviations when found for the same series. For 11 of the 13 characters the Fulton standard deviations are in excess of the average

values and two of the differences in these cases are probably significant. The other two, for which the position is reversed, are quite insignificant. The total series must hence be supposed racially heterogeneous. Standard deviations were also computed for the two later series (mounds 14, 34, 85 and 86) alone, but heterogeneity is still indicated, as 10 of these values out of the 13 exceed the average values for the 14 series.

In spite of the unsatisfactory nature of the total Fulton series, it was thought worth while computing a few coeffieients of racial likeness between it and some other series of American Indian crania used in this paper. A comparison of a few means showed at once that nearly all of the 16 would give reduced values greater than 19 and this was confirmed in four doubtful cases leaving only one below the limit, viz. Fulton ($\bar{n} = 50{\cdot}6$) with Algonkin East-Central (61·9), reduced C.R.L. $= 4{\cdot}42 \pm 0{\cdot}49$ for 12 characters. The fact that this close resemblance is found between two series from the same region is obviously suggestive, but the inadequacy of the new data must not be forgotten. It is to be hoped that the data for additional skulls from Fulton County which are said to be available will make it possible to determine the relationships of the populations represented in a more satisfactory way.

REFERENCES

CLEAVER, F. H. (1937). "A contribution to the biometric study of the human mandible." *Biometrika*, **29**, 80–112.

COLE, FAY-COOPER & DEUEL, THORNE (1937). *Rediscovering Illinois: Archaeological Explorations in and around Fulton County.* Chicago. For crania see App. IV: "Preliminary notes on the crania from Fulton County, Illinois", by Georg K. Neumann.

HRDLIČKA (1919). "Anthropometry D. Skeletal parts: the skull." *Amer. J. Phys. Anthrop.* **2**, 401–28. Reprinted in author's *Anthropometry* (1920). Philadelphia. English trans. of Monaco report given in each.

KITSON, E. (1931). "A study of the Negro skull with special reference to the crania from Kenya Colony." *Biometrika*, **23**, 271–314.

MACCURDY, G. G. (1923). "Human skeletal remains from the highlands of Peru." *Amer. J. Phys. Anthrop.* **6**, 218–329.

MORANT, G. M. (1925). "A study of Egyptian craniology from prehistoric to Roman times." *Biometrika*, **17**, 1–52.

—— (1927). "A study of the Australian and Tasmanian skulls based on previously published measurements." *Biometrika*, **19**, 417–40.

—— (1928). "A preliminary classification of European races based on cranial measurements." *Biometrika*, **20B**, 301–75.

—— (1937). "A contribution to Eskimo craniology based on previously published measurements." *Biometrika*, **29**, 1–20.

MORTON, SAMUEL GEORGE (1839). *Crania Americana; or, a Comparative View of the Skulls of various American Nations of North and South America, to which is prefixed an Essay on the Varieties of the Human Species.* Philadelphia.

PEARSON, KARL & DAVIN, ADELAIDE G. (1924). "On the biometric constants of the human skull." *Biometrika*, **16**, 328–63.

VIRCHOW, RUDOLF (1892). *Crania Ethnica Americana.* Publ. as supplement to *Z. Ethn.*

VON EICKSTEDT (1934). *Rassenkunde und Rassengeschichte der Menschheit.* Stuttgart.

WOO, T. L. & MORANT, G. M. (1932). "A preliminary classification of Asiatic races based on cranial measurements." *Biometrika*, **24**, 108–34.

A Description of Nine Human Skulls from Iran excavated by Sir Aurel Stein, K.C.I.E.

By G. M. MORANT

There are few parts of the world which are less known from a craniological point of view than Iran. The total number of skulls from the country preserved in European collections appears to be less than fifty and no measurements for a series of any length have been published.* The specimens described below were obtained by Sir Aurel Stein during two of his archaeological expeditions, and I am indebted to him for granting me permission to examine them. The material is not extensive enough to justify any statistical comparisons of the measurements, and the object of this note is to place on record particulars of the provenance of the skulls and a description of their characters.

Nos. 1 and 2 were excavated by Sir Aurel Stein on his Third Persian Expedition (1933–4) and the remaining seven on his Fourth Persian Expedition (1936–7). A published account of the discovery of the first skull is quoted below and particulars of the others are from his unpublished records. The condition of the specimens can be seen from the photographs (Plates I–III). Dehbid is in the province of Fars and it may be said to belong to Central Iran, Bampur is the south-east of the country (Persian Baluchistan) and Dinkha and Hasanlu are in the extreme north-west, near Lake Urumiyeh. The sites are thus widely separated except the last two which are about 50 miles apart.

(1) Skull of an infant found in an artificial mound at Dehbid.

"On excavation the mound yielded throughout abundant painted potsherds, worked stones and associated objects from the chalcolithic period of occupation. These were found at depths 1 to 4 ft. below the surface level. In section iii at a depth of 1 ft. were discovered the remains of a partial burial, comprising the neatly trepanned skull of a woman or child, a lower jaw and a small quantity of bone fragments lying close to it. A small carved stone pendant representing a clenched hand subsequently turned up in the same section and depth, but a little farther off. This resembles so closely a number of similar pendants found in one of the Sasanian burial cairns of Bishezard that a strong probability suggests itself of the partial burial near which it was found being intrusive, i.e. having been placed within the chalcolithic debris layer at the foot of the mound in historical times." (Stein, 1936.)

The excellent and fresh condition of the cranial bones renders it extremely probable that the burial was intrusive, and that it belongs not only to historical but also to very recent times. The child probably died in the third year of life. The milk dentition was completely erupted and the crypts for the first and second permanent molars were open in both jaws, with the crowns of these teeth formed but not erupted as far as the alveolar margins. The basi-occipital and right exoccipital bones are missing, while the suture between the left exoccipital and the supra-occipital is open except for a length of 1 cm. where it is synostosed. The hole in the left parietal (see Plate III B) was almost certainly made after death by a blow from a pointed weapon or tool. Its edge shows no sign of separation and the rondelle of bone forced out is still attached to the endocranial surface.

(2) Skull of an adult male from Bampur.

"The skull marked 'Bampur B + 5 feet' was found in a grave on the top of a prehistoric mound near Bampur fort in Persian Baluchistan. It is in all probability mediaeval and

* The longest published series appears to be that of eleven skulls for which measurements and descriptions were provided by the late Dr Viktor Lebzelter (1931).

may have been that of some Baluch belonging to the same tribe as now forms the population of that territory."

This well-preserved skull has clear male characteristics. The coronal suture is just beginning to close. The teeth are considerably worn, eight had been lost before death and three *in situ* are reduced to stumps owing to caries. The upper left canine was formed but unerupted, its tip being on a level with the alveolar margin.

(3) Cranium of an infant from Dinkha.

"The skull from Dinkha was found in a tomb excavated by the eroded side of a large mound occupied in chalcolithic times. The site lies in the large valley of Ushnu, between the south-west shore of Lake Urumiyeh and the main Zagros range forming the boundary between Persian and Iraq Kurdistan."

The specimen consists of a calvaria with the base partly defective and the greater part of the right side of the upper facial skeleton. The child probably died in the third year of life. Judging from the right side of the upper jaw, the milk dentition was completely erupted and the crypts for the first permanent molars were open, with the crowns of these teeth formed but not erupted as far as the alveolar margin. The basi-occipital and right exoccipital bones are missing, and the suture between the left exoccipital and the supra-occipital is half obliterated.

(4) Calvaria of a child from Hasanlu: Hasan. A.

"The six skulls from Hasanlu came from burial of a late chalcolithic period found in an extensive ancient graveyard adjacent to a very large mound near Hasanlu village, some 6 miles to the south of the southern shore of Lake Urumiyeh. The burials comprised complete bodies, all parts except the skulls being much injured. The dead had been buried at depths varying from about 8 to 12 feet. The furniture, mainly pottery, was fairly uniform."

The bones of specimen A from this cemetery are remarkably fresh. The basal suture is completely open and an age at death between 5 and 10 years is suggested by the form and size of the calvaria.

(5) Skull of an adult female from Hasanlu: Hasan. B.

The coronal suture is beginning to close while the sagittal and lambdoid are open. The greater part of the vault was affected by a pathological condition, the ectocranial surface being rugose and in places exceptionally thin, especially at the obelion. Within the area affected the sutures (including the whole of the sagittal suture) are far simpler than usual (see Plate III D). The vault is asymmetrical, the right side being higher than the left (see Plate II B). There is fronto-temporal articulation on both sides. The two upper central incisors were the only teeth lost before death and there is a large abscess cavity at the site of the right tooth (see Plate II B). The upper left canine is small and peg-shaped and one premolar is reduced to a stump owing to caries. The teeth are considerably worn.

(6) Skull of an adult from Hasanlu: Hasan. C.

In spite of its large size, this specimen is probably female, the superciliary ridges and transverse occipital lines being feebly developed. The calvarial sutures are completely open. The right central incisor was the only tooth lost from the upper jaw before death. A premolar and a molar had been lost from the lower jaw and no third molars had erupted. The teeth in both jaws are considerably worn and three had been reduced to stumps owing to caries. There is a large abscess cavity at the socket for the root of the upper right lateral incisor (see Plate II D). The mandible is exceptionally small and feeble for the cranium.

(7) Skull of an adult male from Hasanlu: Hasan. D.

This is a well-developed and muscular specimen. The calvarial sutures are completely open. No teeth had been lost before death and the upper left third molar is absent. One upper molar is markedly eroded by caries and the teeth are moderately worn.

(8) Cranium of an adult female from Hasanlu: Ḣasan. F.

The lambdoid suture is closing, the sagittal is beginning to close and the coronal is open. No teeth had been lost from the upper jaw before death. The teeth are considerably worn. The right side of the palate has a rugose surface and two cavities due to disease.

(9) Calotte of an adult female from Hasanlu: Hasan. G.

The coronal and sagittal sutures are beginning to close on the external surface and nearly obliterated on the internal; the lambdoid is beginning to close on the external and half obliterated on the internal surface.

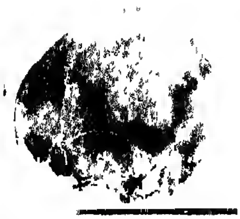
It may be noted that all of the five adult specimens with one or both jaws extant exhibit some form of dental disease, while the age at death for the oldest of these people was probably under 35 years. This would not have been anticipated as the teeth and palates of late prehistoric skulls are usually found to be better preserved than those of modern man. Judging from a qualitative comparison, and making allowances for age and sex, eight of the total nine specimens do not show greater differences than those which might well be found in a sample of such a size selected from a racially homogeneous population. The remaining skull is the modern one from Bampur in Persian Baluchistan and it appears to be distinguished from the others chiefly by the form of its facial skeleton, though it also has the highest cephalic index.

Measurements are provided in Tables I and II, the usual biometric symbols denoting these being given and also the numbers in Rudolf Martin's list. There is nothing particularly remarkable in these data, but it may be noted that if the specimens are considered as a single series the type is decidedly orthognathous. The photographs reproduced in Plates I–III were all taken as nearly as possible with the focal plane of the camera parallel or perpendicular to the Frankfort horizontal plane.

TABLE I

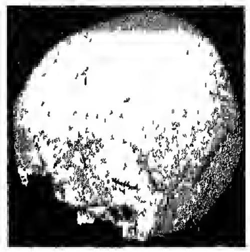
Calvarial measurements of Iranian skulls

	Hasan. A Juv.	Dinhka Juv.	Dehbid Juv.	Hasan. B ♀	Hasan. C ♀?	Hasan. F ♀	Hasan. G ♀	Hasan. D ♂	Bampur ♂
Glabella-occipital max. length (L: M.1)	168 5	152·5	157	175	192·5	174	180	189	164·5
Max. parietal breadth (B. M 8)	130	122	123 5	132·5	133	126	131	140	139
Min frontal breadth (B'. M 9)	92·7	72·3	83·9	93·0	94·1	86·8	—	96·1	92·8
Max. frontal breadth (B'': M.10)	113·5	94	102	112·5?	112·5	103	113	116·5	118 5
Biasterionic breadth (M.12)	95·5	92	96	105	103	103	106 5	109	103·5
Basio-bregmatic height (H', M 17)	123	—	—	128	133	126·5	—	139	133·5
Chord nasion to bregma (S_1': M.29)	109·0	—	93 8	108·7	119·6	100·0	109·4	113·0	104 6
Chord bregma to lambda (S_2': M.30)	110·6	95·8	103·7	112·2?	116·2	115·9	120 2	120·1	107·0
Chord lambda to opisthion (S_3', M.31)	92·8	83·1	91·3	91·1?	102·3	93·5	—	94·0?	93·3
Arc nasion to bregma (S_1. M.26)	124	—	106	124	138	112	128	125·5	116
Arc bregma to lambda (S_2. M 27)	127 5	108·5	118·5	123?	130	133	133·5	131	124·5
Arc lambda to opisthion (S_3: M.28)	106·5	103·5	109·5	112?	127	110	—	115?	105·5
Arc nasion to opisthion (S: M.25)	358	—	333	359	395	355	—	372	346
Horizontal circumference (U · M.23a)	478	427	446	500	532	485	—	523?	479
[illegible] bregma ($\beta Q'$ M.24)	298	262	291	295	312	281	—	319	312
[illegible] *mum* (fml: M.7)	32·0	—	—	36 0?	36 1	37·0	—	40·4?	35·0
Breadth of foramen magnum (fmb M.16)	27·5	—	—	28·0?	25 9?	27 0	—	—	28·0
Chord nasion to basion (LB. M 5)	87·3	—	—	91·8	100·0	94 8	—	110·2	99·9
100 B/L	77·2	80·0	78·7	75·7	69 1	72·4	72·8	74 1	84·5
100 H'/L	73·0	—	—	73·1	69·1	72·7	—	73·5	81·2
100 B/H'	105 7	—	—	103 5	100·0	99 6	—	100 7	104·1
Occipital index (Pearson's)	65·3	57·3	60·1	58·1?	58 5	62·1	—	58 4	67·8
100 fmb/fml	85 9	—	—	77·8?	71·7?	73·0	—	—	80·0

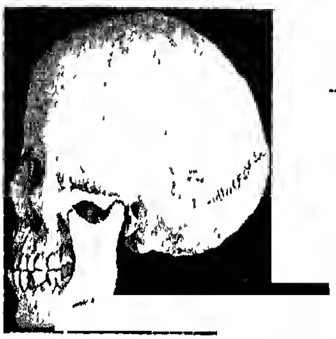


A. Dinkha, juv.

B. Dehbid, juv.



C. Hasan A, juv.



D. Hasan B, ♀

E. Hasan. G, ♀

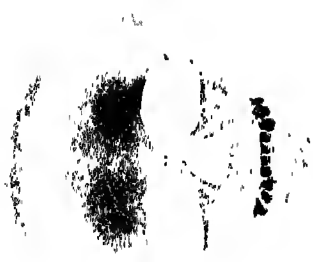
F. Hasan. F, ♀

G. Hasan C, ♀?

H. Bampur, ♂

I. Hasan. D, ♂

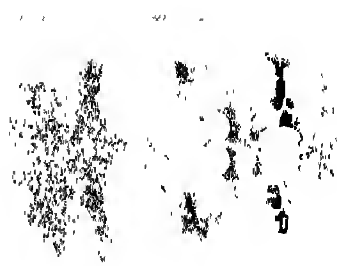
Iranian skulls



A. Dehbid, juv.

B. Hasan. B, ♀

C Hasan. F, ♀



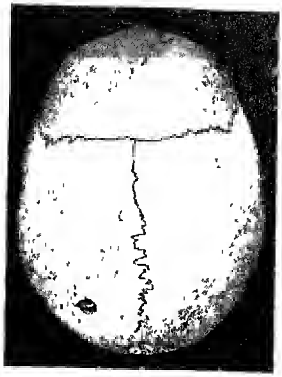
D. Hasan C, ♀?

E. Bampur, ♂

F. Hasan D, ♂

Iranian skulls

A. Hasan, A, juv.



B. Dehbid, juv

C. Dinkha, juv.

D. Hasan, F, ♀

E. Hasan, G, ♀

F. Hasan, B, ♀

G. Hasan, C, ♀?

H. Bampur, ♂

I. Hasan, D, ♂

Iranian skulls

TABLE II

Facial measurements of Iranian skulls

	Dehbid Juv.	Hasan. B ♀	Hasan. C ♀?	Hasan. F ♀	Hasan. D ♂	Bam-pur ♂
Bizygomatic breadth (J: M.45)	96	120 5	—	115	128	128·5
Mid-facial breadth (GB: M.46)	70·6	98·7	93·9	87·8	98 2	96 1
Upper facial height ($G'H$: M.48)	44·2	64·9?	69·0?	59·4	80·9	65 0?
Chord basion to alveolar point (GL)	—	84·0?	86 9?	89 0	106·1	96·4?
Nasal height (NH, L)	31·2	52·1	53·5	42·6	60·8	46·2
Nasal breadth (NB: M 54)	19·5	24·1?	25·2	23·0	24·1	26·0
Orbital breadth L (O_1L: M.51)	33·4	40·9?	38·6	37·7	45·0	43·0
Orbital height L (O_2L: M.52)	27·2	33·6	31·5	31 7	35·4	32·7
Palatal length (G_1': M.62)	33·4	—	44·0	40·5	51·6	45·0
Palatal breadth (G_2: M.63)	—	43·9	43·0	39·3	44·4	—
Simotic chord (SC: M.57)	8·8?	—	9·4	9·0	11 8	11·3
Subtense to simotic chord (SS)	—	—	6 7	2·9	7·4	4·6
100 $G'H/GB$	62·6	65·8?	73·5?	67·7	82·4	67·6?
100 NB/NH, L	62·5	46·3?	47·1	54·0	39·6	56 3
100 O_2/O_1, L	81·4	82 2?	81 6	84·1	78·7	76·0
100 G_2/G_1'	—	—	97·7	97·0	86·0	—
100 SS/SC	—	—	71·3	32·2	62·7	40·7
$N\angle$	—	61°·9?	58°·4?	65°·9	65°·2	67°·7?
$A\angle$	—	75°·1?	79°·0?	76°·8	70°·9	73°·8?
$B\angle$	—	43°·0?	42°·6?	37°·3	43°·9	38°·5?
Alveolar profile angle ($P\angle$)	88°·5	—	92°·5?	83°·5	86°	81°

REFERENCES

LEBZELTER, V. (1931). "Schadel aus Persien." *Ann. naturh. Mus. Wien*, **45**, 137–57.
STEIN, A. (1936). "An archaeological tour in the Ancient Persis." *Iraq*, **3**, 217.

THE PROBABILITY INTEGRAL TRANSFORMATION FOR TESTING GOODNESS OF FIT AND COMBINING INDEPENDENT TESTS OF SIGNIFICANCE

By E. S. PEARSON

1. Introductory

If $p(x)$ is the elementary probability law of a continuous random variable x in the interval $a \leqslant x \leqslant b$, so that $p(x) = 0$ for $x < a$ or $> b$ and

$$\int_a^b p(x)\,dx = 1, \qquad \ldots\ldots(1)$$

then we may write

$$y = \int_a^x p(x)\,dx. \qquad \ldots\ldots(2)$$

y is a non-decreasing function of x, having values confined to the interval (0, 1). Further

$$p(y) = p(x)\Big/\frac{dy}{dx} = 1 \quad \text{for} \quad 0 \leqslant y \leqslant 1. \qquad \ldots\ldots(3)$$

In other words the probability law for the integral, y, is rectangular, all values of y between 0 and 1 being equally likely to occur. It follows that if we wish to use a set of n independent observations $x_1, x_2, \ldots, x_n$ to test the hypothesis H_0 that a probability law is of specified form, say $p(x \mid H_0)$, it may be possible to carry out this by testing the equivalent hypothesis, h_0, that the corresponding values $y_1, y_2, \ldots, y_n$, obtained by means of the transformation (2), have been randomly drawn from the rectangular distribution (3). The relation between x_i and y_i is illustrated in Fig. 1; corresponding to the abscissae x_i, $(i = 1, 2, \ldots, 10)$, of the ten ordinates drawn above, are ten values of y shown below on the scale 0 to 1. The hypothesis H_0 that the ten x's are a random sample from a population distribution represented by the frequency curve is therefore equivalent to the hypothesis h_0 that the ten y's form a random sample from a rectangular distribution, range 0 to 1.

If the probability laws $p(x)$ are not the same for all the x's, so that

$$y_i = \int_{a_i}^{x_i} p_i(x)\,dx \quad (i = 1, 2, \ldots, n), \qquad \ldots\ldots(4)$$

the n values of y_i will still be distributed independently as in (3). It follows that the transformation is applicable not only to problems generally classed under the heading of tests of goodness of fit, where $p_i(x)$ is the same for all i, but also in another important type of problem where x_i are a number of independent test

criteria, e.g. a number of values of "Student's" t or Fisher's z associated with differing degrees of freedom, and it is wished to obtain a single test of a comprehensive hypothesis. Thus for example we may either:

(*a*) Test whether it is likely that a sample of ten values of a variable x has been drawn from a Normal distribution with specified mean and standard deviation, ξ_0 and σ_0.

(*b*) Test the hypothesis that there is no difference between the gain in weight of children fed on (i) raw, (ii) pasteurized milk, using ten values of t obtained from a comparison in ten age groups of the mean difference in weight increase of children fed for six months on the two diets.

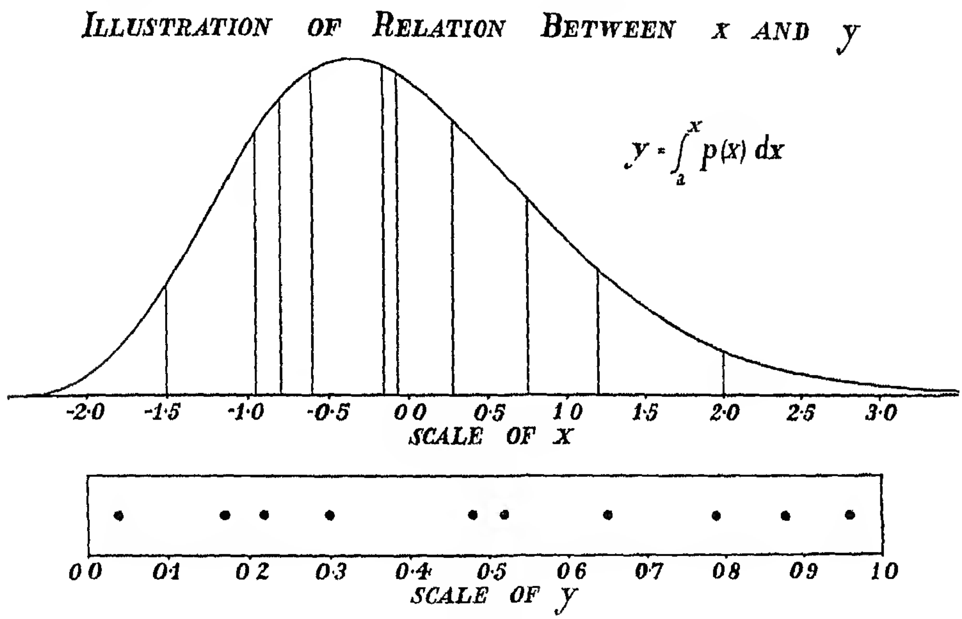


Fig. 1.

Results following from this idea of using the probability integral transformation, which seems likely to be one of the most fruitful conceptions introduced into statistical theory during the last few years, have been developed by R. A. Fisher (1932), Karl Pearson (1933, 1934) and J. Neyman (1937). It is my purpose in this article to review and link together some of the suggestions that have been put forward.

2. CHOICE OF THE APPROPRIATE TEST CRITERION

The probability that in a random sample of size n from the rectangular distribution (3), the y's will fall within the elementary intervals $y_i \pm \frac{1}{2}dy_i$ $(i = 1, 2, \ldots, n)$ is $dy_1 dy_2 \ldots dy_n$, i.e. is independent of the particular values of y. Thus any set of values of y is as likely to occur as another. What criterion are we therefore to

use in testing the hypothesis, h_0, that the sample has been drawn from the rectangular population? Established custom in analogous problems might suggest that we should compare the moments of the sample with those of the rectangle. But which moments and how many? Fig. 2 shows six possible y-samples of size $n = 10$; of these sample (a) is likely to have moments agreeing most closely with those of the rectangle. Nevertheless each of the spot patterns illustrated is equally likely to occur in sampling if h_0 is true, and to assume that the test must be based on moments would appear to prejudice the issue.

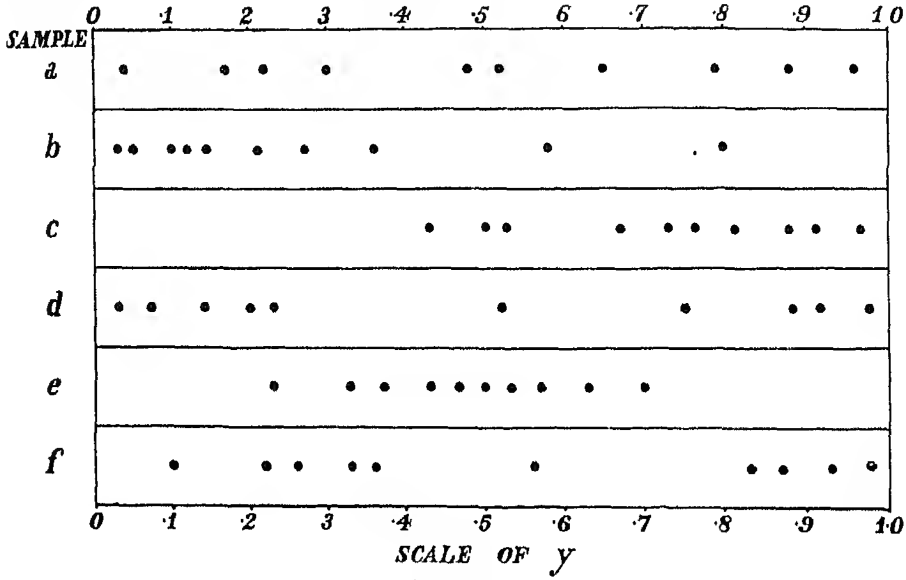


Fig. 2.

Following what may be described as the intuitional line of approach, K. Pearson (1933)* suggested as suitable test criteria one or other of the products

$$Q_1 = y_1 y_2 \cdots y_n, \qquad \ldots\ldots(5)$$

or

$$Q_1' = (1-y_1)(1-y_2)\ldots(1-y_n).\dagger \qquad \ldots\ldots(6)$$

Here Q_1 is the joint probability that in random sampling from $p_i(x)$ the n values of x will be as small or smaller than the corresponding observed values; Q_1' is the probability that they will be as great or greater than their observed values. In Fig 2, sample (b) will give a relatively low value to Q_1, and a relatively high value to Q_1'; for sample (c) the position is reversed. To form a complete statistical test it is clearly necessary to know how these Q criteria are distributed in random sampling if the hypothesis h_0 regarding the y's, and therefore the hypothesis H_0 regarding the x's, were true.

* R. A. Fisher (1932) was primarily concerned with a combination of tests of significance, where the distinction between Q_1 and Q_1' did not arise in the same way.

† K. Pearson denoted these products by λ_n.

By means of a simple transformation to new variables

$$v_i = -2\log_e y_i \quad (i = 1, 2, \ldots, n), \qquad \ldots\ldots(7)$$

it is easy to show that $-2\log_e Q$ is distributed as χ^2 with degrees of freedom $f = 2n$, i.e.

$$p(\chi^2) = \frac{1}{\Gamma(\frac{1}{2}f)\,2^{\frac{1}{2}f}}(\chi^2)^{\frac{1}{2}f-1}e^{-\frac{1}{2}\chi^2}. \qquad \ldots\ldots(8)$$

Exceptionally small values of Q_1 or Q_1' correspond to large values of χ^2. Thus a straightforward test is available which, on choice of the appropriate probability level from the χ^2 tables, gives a precise control of the risk of rejecting the hypothesis tested regarding the $p_i(x)$ when it is true.

In discussing the application of this test K. Pearson was aware of the difficulty of choice between Q_1 and Q_1'. From which tail of the distributions should the probability integral be calculated? He suggested that the smaller of the two should be used as giving the "more stringent test". It may be noted that as an alternative to Q_1 and Q_1' a third criterion may be used, namely

$$Q_2 = \prod_{i=1}^{n}(y_i'), \qquad \ldots\ldots(9)$$

where

$$\left.\begin{aligned} y_i' &= 2\int_a^{x_i} p_i(x)\,dx = 2y_i \text{ if } x_i \text{ is below median } x,\\ &= 2\int_{x_i}^{b} p_i(x)\,dx = 2(1-y_i) \text{ if } x_i \text{ is above median } x. \end{aligned}\right\} \qquad \ldots\ldots(10)$$

It is seen that y_i' follows the rectangular distribution (3) if H_0 is true, and therefore $-2\log_e Q_2$ is also distributed as χ^2 with $f = 2n$. The criterion Q_2 will be exceptionally small if the x's lie towards either tail of their probability distributions,* e.g. in sample (*d*) of Fig. 2; it will be exceptionally large for sample (*e*).

Provided that the test based on one of the products Q is being used to combine together a number of independent tests of significance, the intuition which lead to its choice appears on the whole to be sound, though it cannot be claimed that it is necessarily the best test. In such a problem the separate test criteria x_i (whether t, z, r, χ^2, etc.) have been chosen so that small values of y_i or of $1-y_i$ suggest that the individual hypotheses are improbable. Consequently a small value of Q is essentially associated with improbability of the combined result. Nor will it generally be difficult to decide on *a priori* grounds which of the three forms of Q is appropriate.† In the case of tests of goodness of fit, however, when it is wished to test whether a sample $x_1, x_2, \ldots, x_n$ can have been randomly drawn from a population with probability law $p(x) = p(x \mid H_0)$, there appear to be no *a priori* reasons for choosing the Q type of criterion based on the *product* of the

* This form of the criterion appears first to have been defined precisely in print by P. V. Sukhatme (1935, p. 587).

† It is of course important not to make the decision as to which end of the x-distribution to start from in taking the integral depend on the observed values of the x's.

probability integrals. When all the forms of pattern of the y's, as shown in Fig. 2, are equally likely to occur if h_0 be true, how, it must be asked, are we to settle when the hypothesis should be rejected? It seems only possible to proceed further by specifying what other forms of probability law are to be regarded as possible alternatives to $p(x \mid H_0)$.

Denote by $p(x \mid H_1)$ some alternative law. If now this is the true probability law, but the y's have been calculated from equations (2) on the assumption that $p(x) = p(x \mid H_0)$, then, as Neyman has pointed out,

$$p(y \mid h_1) = \frac{p(x \mid H_1)}{\dfrac{dy}{dx}} = \left.\frac{p(x \mid H_1)}{p(x \mid H_0)}\right|_{x=f(y)} \text{for } 0 \leqslant y \leqslant 1, \qquad \ldots\ldots(11)$$

where $f(y)$ means the solution of

$$y = \int_a^x p(x \mid H_0)\, dx \qquad \ldots\ldots(12)$$

with regard to x. Thus the probability distribution of y, when H_1 is true, is obtained by calculating at points $x = f(y)$ the ratio of the ordinates of the true and hypothetical probability functions. As an example, suppose that we are using n values of x to test the hypothesis that the sampled population is represented by a normal curve with mean at zero and unit standard deviation. Then

$$p(x \mid H_0) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}x^2}. \qquad \ldots\ldots(13)$$

Consider what would be the equation of $p(y \mid h_1)$ if the following had been the true forms of the population sampled:

(I) $$p(x \mid H_1) = \frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-\frac{1}{2})^2}, \qquad \ldots\ldots(14)$$

a normal curve with mean at $+0{\cdot}5$ and unit standard deviation.

(II) $$p(x \mid H_1) = \frac{2}{3\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{2x}{3}\right)^2}, \qquad \ldots\ldots(15)$$

(III) $$p(x \mid H_1) = \frac{3}{2\sqrt{(2\pi)}} e^{-\frac{1}{2}\left(\frac{3x}{2}\right)^2}, \qquad \ldots\ldots(16)$$

normal curves with means at zero and standard deviations of $\frac{3}{2}$ and $\frac{2}{3}$ respectively.

(IV) $$p(x \mid H_1) = c(1 + \tfrac{1}{2} x\sqrt{\beta_1})^{\frac{4}{\beta_1}-1} e^{-\frac{2x}{\sqrt{\beta_1}}}$$

where $\quad (a)\ \sqrt{\beta_1} = 0{\cdot}4, \quad (b)\ \sqrt{\beta_1} = 0{\cdot}7, \qquad \ldots\ldots(17)$

Pearson Type III curves with mean at zero, unit standard deviation and $\beta_1 = 0{\cdot}16$ or $0{\cdot}49$.

Values for $p(y \mid h_1)$ were calculated from (11), corresponding to the points

$y = 0, 0{\cdot}05, 0{\cdot}10, 0{\cdot}20, \ldots, 0{\cdot}80, 0{\cdot}90, 0{\cdot}95, 1{\cdot}00$;* the resulting curves are drawn in Fig. 3. They represent a number of different forms of departure from the rectangular y-distribution, corresponding in $p(x)$ to: (I) a shift in mean; (II) and (III) changes in standard deviation; (IV) a change in shape. Clearly, in Fig. 2,

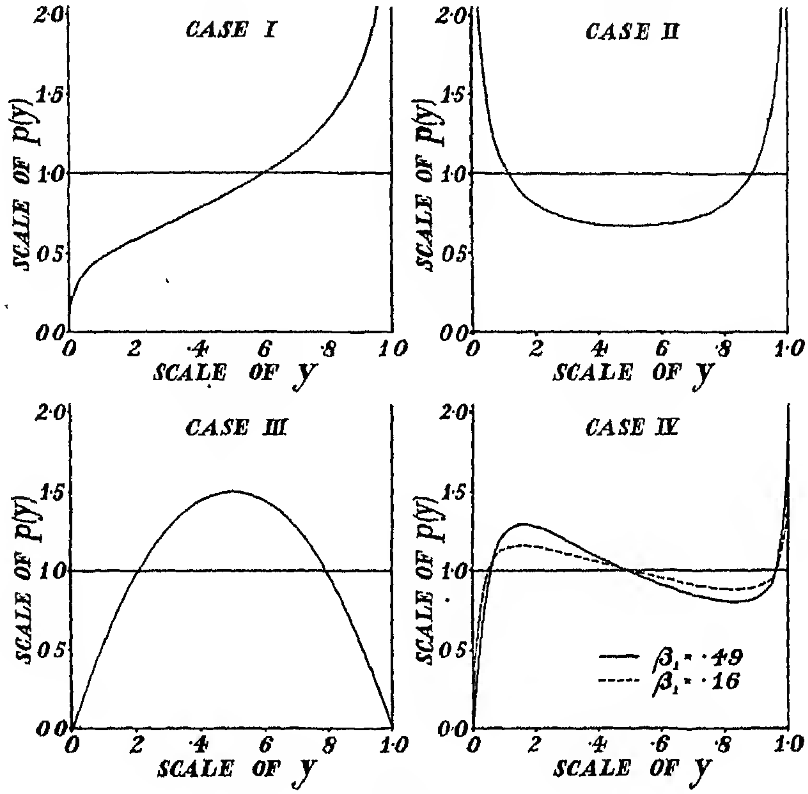


Fig. 3.

Alternatives to $p(x|H_0)=\dfrac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2}$.

I. $p(x|H_1)=\dfrac{1}{\sqrt{(2\pi)}}e^{-\frac{1}{2}(x-\frac{1}{2})^2}$.

II. $p(x|H_1)=\dfrac{2}{3\sqrt{(2\pi)}}e^{-\frac{1}{2}\left(\frac{2x}{3}\right)^2}$.

III. $p(x|H_1)=\dfrac{3}{2\sqrt{(2\pi)}}e^{-\frac{1}{2}\left(\frac{3x}{2}\right)^2}$.

IV. $p(x|H_1)=c(1+\frac{1}{2}x\sqrt{\beta_1})^{\frac{4}{\beta_1}-1}e^{-\frac{2x}{\sqrt{\beta_1}}}$.

samples (c), (d), (e) and (f) are of patterns we might expect to find when testing H_0, if the populations sampled differed from (13) in the directions of (14), (15), (16) and (17) respectively.

The questions, therefore, that need consideration appear to be the following.

* For the Type III curve, the tables of ordinates entered against a standardized abscissa, published by L. R. Salvosa (1930), were found very useful.

In testing, on observed sample values, whether $p(x \mid H_0)$ represents the population probability law,

(i) Can we define in what way the true probability law may diverge from that specified by H_0 (e.g. in location, scaling, shape, etc., one or all)?

(ii) If this is possible, can we determine the most efficient test to apply to the y's in order to detect such divergence if it exists?

(iii) If the definition required in (i) is impossible, how far can we determine what may be called a useful "omnibus" test, sensitive as far as possible to many forms of divergence?

It should be noted, and this point must be emphasized, that it is fundamental to any procedure that we may base on the distribution of y that the n transformed observations $y_1, y_2, \ldots, y_n$ are independent. If the function $p(x \mid H_0)$ is obtained by fitting a frequency curve to a set of observed x's, this condition will not be satisfied by the resulting y's. For example, had the curve been fitted by equating the first two moments of the theoretical distribution to those of the observations, types of pattern like those suggested in samples (*b*), (*c*), (*d*) and (*e*) of Fig. 2 would probably be ruled out, and the distribution of $-2\log_e Q$ could no longer be that of χ^2. Whether some method of applying a test to the y's can still be devised under these conditions has yet to be investigated.

It must also be borne in mind that once we admit it to be necessary to take into account the form of the alternative hypotheses, a difference in character appears between the goodness of fit problem and that which is concerned with combining independent tests of significance. In the former case, if H_0 is not true, we suppose there exists some common alternative form $p(x \mid H_1)$ appropriate for all the x's, and hence a common $p(y \mid h_1)$. In the latter case, while the different $p_i(x \mid H_0)$ will lead on transformation to a common $p(y \mid h_0) = 1$, the alternatives $p_i(x \mid H_1)$ will not necessarily lead to a common $p(y \mid h_1)$ for all the test criteria. In the two following sections it is primarily the first type of problem that will be considered; the conclusions reached will, however, throw some light on the position obtaining in the second case.

3. A partial solution based on the product criteria, Q

The curves corresponding to cases I, II and III in Fig. 3 could all be graduated roughly by Pearson Type I curves of the form

$$p(y) = p(y \mid h_1) = \frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)}\, y^{m_1}(1-y)^{m_2}. \qquad \ldots\ldots(18)$$

In the case when the hypothesis tested is true, i.e. $h_1 = h_0$, the rectangular distribution results from setting $m_1 = m_2 = 0$. The curve

$$p(y \mid h_1) = (m+1)(1-y)^m, \quad -1 < m \leqslant 0, \qquad \ldots\ldots(19)$$

while it has an ordinate of value $(m+1) < 1$ at $y = 0$, provides an approximation to the form of the curve in Case I. Again the curve

$$p(y \mid h_1) = \frac{\Gamma(2m+2)}{\{\Gamma(m+1)\}^2} y^m(1-y)^m, \qquad \text{......(20)}$$

can be made to represent the y-distributions of Case II ($m < 0$) and Case III ($m > 0$). No Type I curve can represent the y-distributions in Case IV.

Starting from (18), or its special forms (19) and (20) as representing the possible alternatives, it is of interest to see what criterion, for testing the hypothesis h_0 (that $p(y)$ is rectangular), flows from the application of the likelihood method which J. Neyman and the present writer have made frequent use of in other problems.

This method consists of the following procedure:

(1) Given a sample of n independent observations $y_1, y_2, \ldots, y_n$, their joint elementary probability law if h_0 be true, is,

$$p(y_1, y_2, \ldots, y_n \mid h_0) = 1, \qquad \text{......(21)}$$

while if any other member of the admissible set of alternatives is true, it is

$$p(y_1, y_2, \ldots, y_n \mid h_1) = \left\{\frac{\Gamma(m_1+m_2+2)}{\Gamma(m_1+1)\,\Gamma(m_2+1)}\right\}^n \prod_{i=1}^{n} y_i^{m_1}(1-y_i)^{m_2}. \qquad \text{......(22)}$$

(2) Determine the values of m_1 and m_2 which make (22) a maximum, and call the corresponding maximized function $p(y_1, y_2, \ldots, y_n \mid h\max)$.

(3) Then the likelihood ratio criterion for testing h_0 will be λ, where

$$\lambda = \frac{p(y_1, y_2, \ldots, y_n \mid h_0)}{p(y_1, y_2, \ldots, y_n \mid h\max)}. \qquad \text{......(23)}$$

Taking the form (19) to represent $p(y \mid h_1)$, we have only one parameter, m, to determine

$$\log p(y_1, y_2, \ldots, y_n \mid h_1) = n \log(m+1) + m \log\left\{\prod_{i=1}^{n}(1-y_i)\right\}. \qquad \text{......(24)}$$

Whence
$$\frac{\partial \log p}{\partial m} = \frac{n}{m+1} + \log Q_1',$$

where Q_1' is defined in (6) above. Equating this expression to zero, it is seen that a maximum solution is given by

$$m+1 = \left\{-\frac{1}{n}\log Q_1'\right\}^{-1} = \frac{2n}{\chi^2}, \qquad \text{......(25)}$$

where
$$\chi^2 = -2\log Q_1' \qquad \text{......(26)}$$

provided that $\chi^2 \geqslant 2n$. If $\chi^2 < 2n$, since $m \leqslant 0$, the maximum solution is given by $m = 0$.

Consequently we find that

$$\lambda = \left(\frac{\chi^2}{2n}\right)^n (e^{-\frac{1}{2}\chi^2})^{1-\frac{2n}{\chi^2}}$$

$$= (2n)^{-n}(\chi^2)^n e^{-\frac{1}{2}\chi^2+n}, \qquad \ldots\ldots(27)$$

provided that $\chi^2 \geqslant 2n$; if $\chi^2 < 2n$ then $\lambda = 1$.*

Thus $\lambda \to 0$ and the hypothesis tested becomes less and less likely when $\chi^2 \to \infty$ and $Q_1' \to 0$. If the hypothesis is true, then we know from the discussion on p. 137 above that $-2\log Q_1'$ is distributed in the standard χ^2 form with $2n$ degrees of freedom.

But not only is the test based on Q_1' that derived from the λ-criterion; it may be easily shown that it is the uniformly most powerful test† of the hypothesis h_0 with regard to the set of alternatives defined by (19). In other words if the admissible alternatives to H_0 lead to forms $p(y \mid h_1)$ following the J-curve (19), then the test based on Q_1' or, if the integrals are more appropriately calculated from the lower terminal, on Q_1 has the following unique property: *it is impossible to find any other test which gives a larger chance of detecting departure of the probability law from the specified form* $p(x \mid H_0)$. A fresh light seems therefore to be thrown on the product criteria Q_1 and Q_1'. While the form (19) will not be exactly followed in practice, a little reflection on the matter suggests that it will represent the general characteristics of the departure of $p(y)$ from the rectangle if the possible changes in $p(x)$ correspond to a translation of the whole $p(x)$ distribution to right (or left). It is of interest to note that apart from its application in goodness of fit tests, this is also the kind of change we may often expect when the x's are the criteria used in a number of independent tests of significance. Thus, if some general hypothesis is not true, a number of independent values of "Student's" t may be distributed approximately about some common mean value other than zero; while the shape and standard deviations of these modified t-distributions will also be altered, the changes involved would relatively be much less than in the mean.

If now we start from the form (20), which as has been pointed out will represent approximately the curves of Cases II and III in Fig. 3, we may proceed to calculate the λ-criterion in a similar manner. The equation to solve for m, to obtain $p(y_1, y_2, \ldots, y_n \mid h\max)$, is

$$\frac{\partial \log \Gamma(2m+2)}{\partial m} - \frac{2\partial \log \Gamma(m+1)}{\partial m} = -\frac{1}{n}\log(Q_1 Q_1'), \qquad \ldots\ldots(28)$$

* Since the admissible alternatives have been restricted to those defined by (19), i.e. with $-1 < m \leqslant 0$, we cannot reject H_0 when high values of Q_1' or low values of χ^2 are obtained from the data. Thus in such cases the value of the λ-criterion is unity, suggesting no reason for rejecting H_0. If however we take $-1 < m < \infty$, equation (19) will now represent J-curves with maxima either at $y=1$ or $y=0$. We are then aiming at a test which is sensitive to translation of $p(x \mid H_1)$ both to right and to left of $p(x \mid H_0)$, and $\lambda \to 0$ either when $Q_1' \to 0$ or when $Q_1' \to 1$.

† See Neyman and Pearson (1933*a*, *b*). The proof that the test possesses this property follows from the results given on pp. 298–302 of the earlier paper.

where
$$Q_1 Q_1' = \prod_{i=1}^{n} y_i(1-y_i). \qquad \ldots\ldots(29)$$

Thus it appears that λ, if determined, would be a function of $Q_1 Q_1'$. Without attempting to go further into the problem it may be noted that a test criterion depending on $Q_1 Q_1'$ is likely to be rather closely correlated with the criterion Q_2, defined in equations (9) and (10). It will be seen that

$$Q_2 = \prod_{i=1}^{n} \{1-2\,|\,y_i - \tfrac{1}{2}\,|\}, \qquad \ldots\ldots(30)$$

and the functions (a) $1-2\,|\,y_i-\frac{1}{2}\,|$, and ($b$) $4y_i(1-y_i)$ both equal zero when $y_i = 0$, increase monotonically to 1 when $y_i = \frac{1}{2}$ and then decrease to 0 as y_i increases to 1. In so far as this correspondence exists, it points to Q_2 being an appropriate criterion when the alternatives to $p(x\,|\,H_0)$ are likely to have the same mean but either larger or smaller standard deviations.

Using the more general Type I form of equation (18), it is found that λ will be a function of both Q_1 and Q_1', but not of $Q_1 Q_1'$.

Finally it must again be noted that (18) cannot represent the curves shown as Case IV in Fig. 3, which arose when the probability distribution $p(x\,|\,H_1)$ had the same mean and standard deviation as $p(x\,|\,H_0)$ but was a skew rather than a normal curve. It is noted however that for both alternatives represented, i.e. Type III curves with $\beta_1 = 0{\cdot}16$ and $0{\cdot}49$ respectively, the gradient of $p(y\,|\,h_1)$ increases approximately from $y = 0$ to $0{\cdot}2$, decreases from $y = 0{\cdot}2$ to $0{\cdot}8$, and increases again from $y = 0{\cdot}8$ to $1{\cdot}0$. Bearing in mind that a criterion of the type $Q_1' = \prod_{i=1}^{n}(1-y_i)$ appears to be efficient in detecting the existence of an increasing gradient as in Case I, the following criterion is tentatively suggested as suitable to detect the presence of skewness:

$$Q_3 = \prod_{i=1}^{n} (y_i'), \qquad \ldots\ldots(31)$$

where
$$\left.\begin{aligned} y_i' &= 5(0{\cdot}2 - y_i) && \text{for } 0 \leqslant y_i \leqslant 0{\cdot}2, \\ y_i' &= \tfrac{5}{3}(y_i - 0{\cdot}2) && \text{for } 0{\cdot}2 < y_i \leqslant 0{\cdot}8, \\ y_i' &= 5(1-y_i) && \text{for } 0{\cdot}8 < y_i \leqslant 1. \end{aligned}\right\} \qquad \ldots\ldots(32)$$

It will be found that if y_i follows the rectangular distribution, so also does y_i'. Thus $-2\log_e Q_3$ will again be distributed as χ^2 with $f = 2n$, if H_0 is true.

The difference in the character of the critical regions of the tests associated with Q_1, Q_2 and Q_3 may be illustrated diagrammatically for the case $n = 2$, where for clearness a 20 % significance level (rather than, say, 0·05 or 0·01) has been taken. In each case the hypothesis H_0 (or h_0) would be rejected if the sample point (y_1, y_2) falls within the shaded regions; if H_0 be true the sample point is equally likely to fall anywhere within the unit square, so that the area of the shaded portions must be 20 % of the whole. The boundaries of these regions were obtained from the χ^2-transformation. Thus for $f = 4$, the upper and lower 20 %

levels for χ^2 are 1·649 and 5·989 respectively, giving corresponding levels for Q of 0·0501 and 0·4385. To determine the boundaries it is then necessary to find the co-ordinates of y_1 and y_2 satisfying, (i) equation (5) for Q_1, (ii) equations (9) and (10) for Q_2; (iii) equations (31) and (32) for Q_3. A sample such as (*b*) of Fig. 2 will give a y-point in the n-dimensioned cube which is likely to fall into the critical

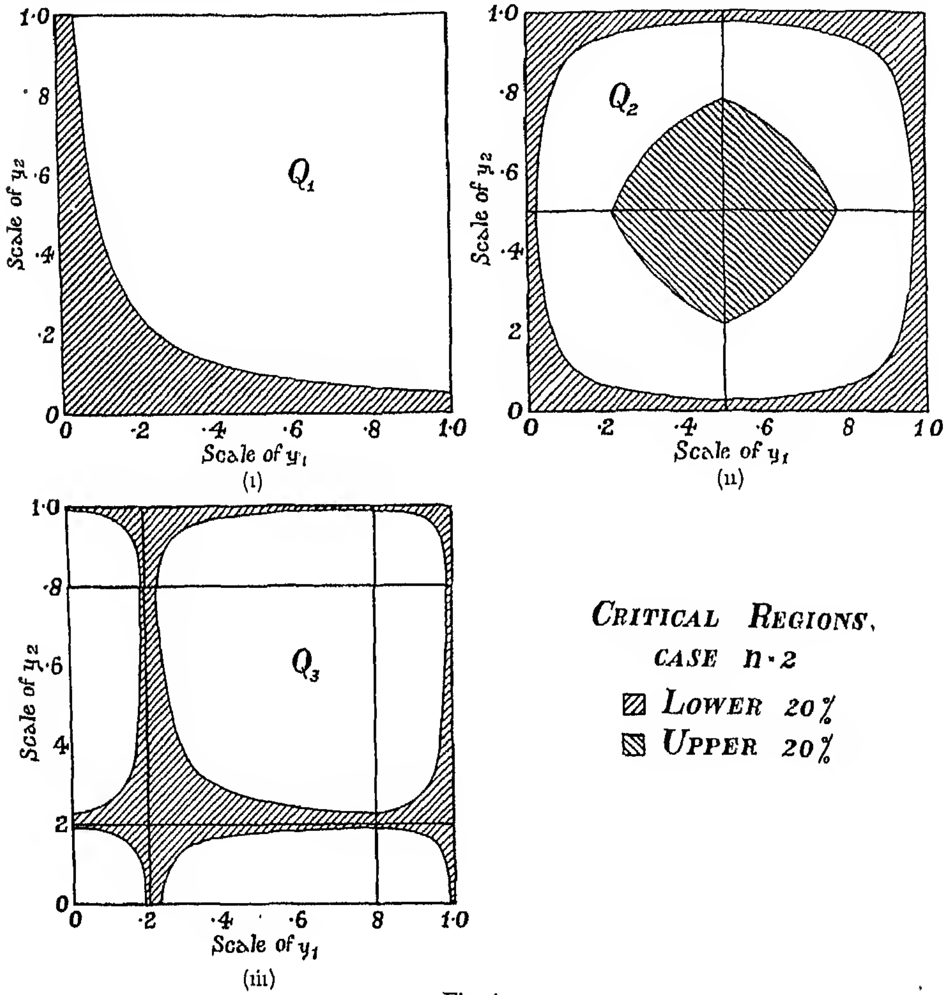


Fig. 4.

region of the type shown in Fig. 4 (i); a sample as (*d*) is likely to give a point falling into a region of the outer ring type of Fig. 4 (ii), while a sample like (*e*) will give a point falling in the central lozenge-shaped type of region of the same diagram. On the other hand samples like (*f*), which seem likely to arise when $p(x \mid H_1)$ has the same mean and standard deviation but greater positive skewness than $p(x \mid H_0)$, will tend to give points in the more complicated region of the type of Fig. 4 (iii), i.e. points with y values between 0·1 and 0·4 or above 0·9.

The suggestion regarding Q_3 is only put forward tentatively. But it appears that in so far as we know the kind of departure from $p(x \mid H_0)$ to be expected, and therefore know the points within the n-dimensioned y-hypercube round which the sample points are likely to cluster, it should be possible to construct appropriate tests of the Q-type on the lines suggested in the case of Q_3. The sampling distribution of such criteria will be always exactly known if H_0 is true,* through the transformation $-2\log_e Q = \chi^2$, while their efficiency in detecting that H_0 is false can be secured on a basis which, if crude, has a definite guiding principle behind it For a more precise handling of the problem Dr Neyman's work on "smooth tests" must be considered.

4. Dr J. Neyman's method of choosing appropriate test criteria

Neyman (1937) deals with the goodness of fit type of problem, that is to say, he supposes that if $p(y)$ is not a rectangle, then some single alternative $p(y \mid h_1)$ is appropriate for all observations. The system of curves which he has taken to represent the possible alternatives is

$$p(y \mid h_1) = p(y \mid \Theta_1, \Theta_2, \ldots, \Theta_k) = ce^{\sum_{i=1}^{k} \Theta_i \pi_i(y)} \quad \text{for } 0 \leqslant y \leqslant 1. \qquad \ldots\ldots(33)$$

These curves depend on k parameters Θ_i which are at our choice; if all the Θ_i's are zero, $p(y \mid h_1) = p(y \mid h_0)$. c is a function of the Θ_i's. Further $\pi_1, \pi_2, \ldots, \pi_k$ are a system of polynomials in y, orthogonal and standardized in the interval (0, 1) of which the first few are as follows.

$$\left.\begin{aligned} \pi_1(y) &= \sqrt{12}(y-\tfrac{1}{2}), \\ \pi_2(y) &= \sqrt{5}\{6(y-\tfrac{1}{2})^2-\tfrac{1}{2}\}, \\ \pi_3(y) &= \sqrt{7}\{20(y-\tfrac{1}{2})^3-3(y-\tfrac{1}{2})\}, \\ \pi_4(y) &= 210(y-\tfrac{1}{2})^4-45(y-\tfrac{1}{2})^2+\tfrac{9}{8}. \end{aligned}\right\} \qquad \ldots\ldots(34)$$

This form for $p(y \mid h_1)$ was chosen by Neyman partly for simplicity in the development of the appropriate tests and partly on the grounds that any function having the characteristics of $\log p(y)$ can be represented by a series of such orthogonal polynomials $\pi_i(y)$. How many and which terms of such a series are needed to represent curves of such varied form as those shown in Fig. 3 has still to be explored. It will be noted that using only $\pi_1(y)$, (33) gives an exponential which will correspond roughly to Case I, Fig. 3. Again $\pi_2(y)$ will lead to a curve that will approximate to Cases II and III, according as Θ_2 is positive or negative, while $\pi_3(y)$ will introduce a point of inflexion of the kind shown for Case IV. Nevertheless it will be seen that the form (33) may need a considerable number of terms before it will make $p(y \mid h_1)$ approach the values of 0 or ∞ at $y = 0$ and 1.

* In this property the tests are more exact than Neyman's tests discussed in the next section, since the sampling distribution of his criteria are only approximate for small values of n.

In some cases therefore the Pearson Type I curve of (18) may be more suitable than (33). It must be remembered, however, that the curves drawn in Fig. 3 are somewhat exceptional, since the differences between the $p(x \mid H_0)$ of (13) on the one hand and the alternatives (14)–(17) on the other are relatively large. In any practical case where n is not too small, one would hope to be able to detect much smaller differences, i.e. to be dealing with alternative distributions $p(y \mid h_1)$ differing less drastically from the rectangular $p(y \mid h_0) = 1$.

Starting from the basis of equation (33), and assuming that n is not too small, Neyman has developed a series of tests, relatively simple to apply, which he calls "smooth tests" that have the following properties.

(*a*) The particular test which is most appropriate will depend upon the number of polynomials needed in (33) to represent the type of departure from the rectangular form likely to be met with in $p(y \mid h_1)$. This is a point at which practical experience must be introduced. Let it be supposed that in a given problem the first k polynomials are regarded as adequate.

(*b*) The test is so adjusted that when H_0 (or h_0) is true, i.e. when $\Theta_1 = \Theta_2 \dots \Theta_k = 0$, the significance level may be fixed at any desired magnitude, e.g. at 0·05 or 0·01.

(*c*) If H_0 be not true, the test is unbiassed in the sense of Neyman and Pearson (1936, 1938), and is more likely than any other unbiassed test to detect departures from zero in the k parameters Θ_t, i.e. to detect that in the place of $p(x \mid H_0)$ some alternative form of law $p(x \mid H_1)$ holds good.

(*d*) The chance of detection, or the power of the test in Neyman and Pearson's terminology, in the neighbourhood of $\Theta_1 = \Theta_2 \dots \Theta_k = 0$ is approximately a function of

$$\phi^2 = \Theta_1^2 + \Theta_2^2 + \dots + \Theta_k^2. \qquad \dots\dots(35)$$

(*e*) For alternatives to $p(x \mid H_0)$ which lead to a function $p(y \mid h_1)$ needing for its representation *more* than the first k-polynomials, the test will not be sensitive. This means that for an "omnibus" test capable of detecting all manner of departures from the rectangle, we may require to introduce a considerable number of polynomial terms. Such a test will however be less efficient in detecting those forms of departure which one or two polynomial terms would be adequate to represent.

If we write

$$z_i = y_i - \tfrac{1}{2} \qquad \dots\dots(36)$$

and

$$\left.\begin{aligned} u_1^2 &= n^{-1}\left\{\sum_{i=1}^{n} \pi_1(y_i)\right\}^2 = 12n^{-1}\left\{\sum_{i=1}^{n} (z_i)\right\}^2, \\ u_2^2 &= n^{-1}\left\{\sum_{i=1}^{n} \pi_2(y_i)\right\}^2 = 180n^{-1}\left\{\sum_{i=1}^{n} (z_i^2) - \tfrac{1}{12}n\right\}^2, \\ u_3^2 &= n^{-1}\left\{\sum_{i=1}^{n} \pi_3(y_i)\right\}^2 = 7n^{-1}\left\{20\sum_{i=1}^{n} (z_i^3) - 3\sum_{i=1}^{n} (z_i)\right\}^2, \end{aligned}\right\} \qquad \dots\dots(37)$$

etc.,

then Neyman's criterion for the kth order test is

$$\psi_k^2 = \sum_{i=1}^{k} (u_i^2), \qquad \text{......(38)}$$

which is approximately distributed as χ^2 with k degrees of freedom. The approximation is due to the fact that while, if h_0 is true, the u's have each an expectation of zero, a unit standard deviation and are *uncorrelated* (i.e. the correlation coefficient between any two of them is zero), they are not independent nor exactly normally distributed. As the sample size, n, increases the accuracy will rapidly improve.

It may be shown that when n is large, and the constant c in (33) assumes its limiting form, $\exp[-\frac{1}{2}\Sigma\Theta_i^2]$, then Neyman's test criterion (38) is exactly that which follows from applying to formula (33) the likelihood method of approach used in the preceding section. In fact it is found that

$$\lambda = e^{-\frac{1}{2}n\sum_{i=1}^{k}(u_i^2)} \qquad \text{......(39)}$$

an expression decreasing from 1 to 0 as the ψ_k^2 of (38) increases from 0 to ∞.

It will be noticed that Neyman's criterion is a sum of polynomial terms in the y_i's, or more simply, using (36), in the z_i's. The product criteria Q_1 and Q_2 of equations (5) and (9) may also be expressed in this form. Thus

$$\begin{aligned}
-\log Q_1 &= -\sum_{i}^{n} \{\log y_i\} \\
&= -\sum_{i}^{n} \{\log(1 + z_i - \tfrac{1}{2})\} \\
&= -\sum_{i} (z_i - \tfrac{1}{2}) + \tfrac{1}{2}\sum_{i} (z_i - \tfrac{1}{2})^2 - \tfrac{1}{3}\sum_{i} (z_i - \tfrac{1}{2})^3 + \ldots, \qquad \text{......(40)} \\
-\log Q_2 &= -\sum_{i} \{\log(1 - 2\,|z_i|)\} \\
&= 2\sum_{i} |z_i| + 2\sum_{i} (z_i^2) + \tfrac{8}{3}\Sigma\,|z_i^3| + \ldots. \qquad \text{......(41)}
\end{aligned}$$

These series do not of course bear any immediate relation to Dr Neyman's polynomial expansions (37).

5. Summary

This paper has drawn attention to the somewhat novel character of the problem to be faced in dealing with tests based on the probability integral transformation. The intuitional notions that have often served to determine the most appropriate test when dealing with normal variation are hardly applicable when we are concerned with a variable following the rectangular distribution. The tests proposed by R. A. Fisher and K. Pearson have been discussed, and emphasis has been laid on the need for consideration of the possible alternatives

to the hypothesis tested. The situation will differ according to whether the problem is one of testing goodness of fit or of combining the results of a number of independent tests of significance. Some illustration of these ideas has been given in the case where the hypothesis regarding the form of a probability law $p(x)$ is incorrect (*a*) in the position of the mean, (*b*) in the magnitude of the standard deviation, (*c*) in the shape of the probability curve. A method has been suggested of adopting the product criteria, Q, to meet these different cases.

Finally, a summary has been given of J. Neyman's suggestions for dealing with the problem. From the theoretical point of view these suggestions appear to be fundamental in character; it is hoped however that it will be possible before long to carry out further numerical investigations (*a*) to determine how large the number of variables, x, must be to make his results accurate for practical purposes; (*b*) to throw more light on the relation between his polynomial form for $p(y \mid h_1)$, the tests based on Q_1, Q_2, Q_3, ..., discussed in preceding sections and the classes of alternatives met with in different types of statistical problem.

REFERENCES

FISHER, R. A. (1932). *Statistical Methods for Research Workers*, 4th ed. § 21·1.
NEYMAN, J. (1937). *Skand. AktuarTidskr.* pp. 149–99.
NEYMAN, J. and PEARSON, E. S. (1933*a*). *Philos. Trans.* A, **231**, 289–337.
—— —— (1933*b*). *Proc. Camb. Phil. Soc.* **29**, 492–510.
—— —— (1936). *Statist. Res. Mem.* **1**, 1–37.
—— —— (1938). *Statist. Res. Mem.* **2** (in the Press).
PEARSON, K. (1933). *Biometrika*, **25**, 379–410.
—— (1934). *Biometrika*, **26**, 425–42.
SALVOSA, L. R. (1930). *Ann. Math. Statist.* **1**, 191–8.
SUKHATME, P. V. (1935). *Proc. Ind. Acad. Sci.* **2**, 584–604.

Note by J. Neyman. I am grateful to the author of the present paper for giving me the opportunity of expressing my regret for having overlooked the two papers by Karl Pearson quoted above. When writing the paper on the "Smooth test for goodness of fit" and discussing previous work in this direction, I quoted only the results of H. Cramér and R. v. Mises, omitting mention of the papers by K. Pearson. The omission is the more to be regretted since my paper was dedicated to the memory of Karl Pearson.

ON TESTS FOR HOMOGENEITY

By B. L. WELCH, Ph.D.

1. Introduction

The present paper is concerned with the familiar E^2* and χ^2 tests for homogeneity. We are given a set of k samples and ask whether they can reasonably be regarded as having all been drawn from one homogeneous population. Denote the samples by $t = 1, 2, \ldots, k$, let there be n_t observations in the tth sample, and denote these by $i = 1, 2, \ldots, n_t$. Then using S to mean $\sum_{t=1}^{k} \sum_{i=1}^{n_t}$ we have

$$E^2 = \frac{S(x_{t.} - x_{..})^2}{S(x_{ti} - x_{..})^2}. \qquad \ldots\ldots(1)$$

A significantly large value of E^2 is taken to denote heterogeneity, levels of significance being deduced from the Beta-function distribution

$$p(E^2) = \frac{1}{B(\frac{1}{2}(k-1), \frac{1}{2}(N-k))} (E^2)^{\frac{1}{2}(k-3)} (1 - E^2)^{\frac{1}{2}(N-k-2)}. \qquad \ldots\ldots(2)$$

This is the distribution which E^2 follows if the k populations sampled are identical and if, in addition, they are normal. In the application of the test in practice we are assuming that departures from normality are not such that (2) is much altered.

It has been argued that the above method of using E^2 may not always be appropriate, depending as it does on the interpretation of the observations as random samples from an infinite hypothetical population. It may sometimes be better to consider the observations as samples from a limited population which is conceived as follows. There are in the aggregate of the k samples $N = \sum_{t=1}^{k} n_t$ observations. These may be divided up into k groups, with n_t in the tth group, in $N!/(n_1! n_2! \ldots n_k!)$ ways. The particular way in which the N observations are grouped in the samples which we are given, may be regarded as randomly selected from all the possible ways of grouping these same N observations. We may calculate a corresponding distribution of values of E^2 (discrete of course) to which the observed E^2 may be referred, instead of (2). This point of view would seem to be particularly appropriate in experimental work where some process of randomization has actually been carried out. For instance there may be k experimental treatments and N experimental objects on which to try them out. If the treatments are assigned to the objects at random, with the sole proviso that each

* E^2 is used instead of η^2 to denote the squared correlation ratio in a sample. This is in accord with the accepted practice of retaining Greek letters for population parameters. χ^2, however, is too well established to be replaced by an italic letter.

treatment shall be repeated a certain number of times, then the connection with the idea of sampling a limited universe is direct.

However, even when the observations are to be interpreted as a sample from an infinite population, it may still be instructive as a first step to consider the limited population which can be generated by shuffling the aggregate of N and redividing into groups of n_t $(t=1,2,\ldots,k)$ in all possible ways. For instance, if we require the moments of E^2 in samples from the infinite population, it may be convenient to calculate them first for the limited population, and then proceed by considering all possible limited populations. This is the procedure of the present paper. The 2-group situation has recently been discussed by E. J. G. Pitman (1937). What follows links up with this work and also has points of contact with an older paper on similar topics by R. C. Geary (1927). I shall also refer to the recent discussions of the χ^2 test for homogeneity when expectations are small, by W. G. Cochran (1936) and J. B. S. Haldane (1937).

2. Sampling a limited population

We shall first derive the mean and variance of E^2 in samples from the limited population. Since the denominator of E^2 is independent of the way in which the aggregate of N is divided into groups, we need only consider the mean and variance of the numerator $S(x_t - x_{..})^2 = S_1$ (say).

When we wish to speak of the observations as an undivided set we shall denote them by y_j $(j=1,2,\ldots,N)$. When we consider the observations divided into k samples, as they are when given to us, we shall denote them, as hitherto, by x_{ti}. The x_{ti} are regarded as a random partition of the y_j into k groups. For the purpose of the present section there is no loss of generality in choosing the origin so that $\sum_j y_j = 0$. (This involves, of course, $\sum_t \sum_i x_{ti} = 0$.) It will be convenient to write as the second and fourth cumulants of the N y's,

$$K_2 = \frac{\left(\sum_j y_j^2\right)}{(N-1)}, \qquad \ldots\ldots(3)$$

$$K_4 = \frac{N(N+1)\left(\sum_j y_j^4\right) - 3(N-1)\left(\sum_j y_j^2\right)^2}{(N-1)(N-2)(N-3)}, \qquad \ldots\ldots(4)^*$$

and to denote expectations over the limited universe by $\mathscr{E}$. Then

$$S_1 = \sum_t \frac{\left(\sum_i x_{ti}\right)^2}{n_t} = \sum_t \frac{\left(\sum_i x_{ti}^2 + \sum_{i \neq i'} x_{ti} x_{ti'}\right)}{n_t}. \qquad \ldots\ldots(5)\dagger$$

* The notation K_2 and K_4 is used instead of R. A. Fisher's k_2 and k_4 to avoid confusion with k, which has been used for the number of groups.

† $\sum_{i \neq i'}$, by this convention, contains $n_t(n_t-1)$ terms, but only $\frac{1}{2}n_t(n_t-1)$ are different.

We note that any term x^2 will have the same expectation, viz.

$$\mathscr{E}(x^2) = \frac{\left(\sum_j y_j^2\right)}{N} = \frac{(N-1)K_2}{N}, \qquad \ldots\ldots(6)$$

and any term of form xx' (i.e. product of two different x's) will have the same expectation, viz.

$$\mathscr{E}(xx') = \mathscr{E}\left(x\frac{(-x)}{(N-1)}\right) = \frac{-\mathscr{E}(x^2)}{(N-1)} = \frac{-K_2}{N}. \qquad \ldots\ldots(7)$$

To obtain the expectation of S_1, it is therefore simply a matter of counting how many terms of each kind there are in (5) and using (6) and (7). We find

$$\mathscr{E}(S_1) = (k-1)K_2. \qquad \ldots\ldots(8)$$

To obtain the variance $V(S_1)$ we have

$$S_1^2 = \sum_t \frac{\left(\sum_i x_{ti}\right)^4}{n_t^2} + \sum_{t \neq t'} \frac{\left(\sum_i x_{ti}\right)^2 \left(\sum_i x_{t'i}\right)^2}{n_t n_{t'}}. \qquad \ldots\ldots(9)$$

This involves terms of five types, viz. x^4, x^3x', $x^2x'^2$, $x^2x'x''$ and $xx'x''x'''$. Each term of a given type has the same expectation. It is sufficient therefore to count how many terms of each type appear in $\left(\sum_i x_{ti}\right)^4$ and $\left(\sum_i x_{ti}\right)^2 \left(\sum_i x_{t'i}\right)^2$. These counts are shown in Table I.

TABLE I

Type of term	No. of terms of each type in $\left(\Sigma_i x_{ti}\right)^4$	No. of terms of each type in $\left(\Sigma_i x_{ti}\right)^2 \left(\Sigma_i x_{t'i}\right)^2$
x^4	n_t	—
x^3x'	$4n_t(n_t-1)$	—
$x^2x'^2$	$3n_t(n_t-1)$	$n_t n_{t'}$
$x^2x'x''$	$6n_t(n_t-1)(n_t-2)$	$n_t n_{t'}(n_t+n_{t'}-2)$
$xx'x''x'''$	$n_t(n_t-1)(n_t-2)(n_t-3)$	$n_t n_t(n_t-1)(n_{t'}-1)$

Making the necessary summations over t, (9) gives

$$\mathscr{E}(S_1^2) = \left(\sum_t \frac{1}{n_t}\right)\mathscr{E}(x^4) + 4\left(k - \sum_t \frac{1}{n_t}\right)\mathscr{E}(x^3x') + \left(k^2 + 2k - 3\sum_t \frac{1}{n_t}\right)\mathscr{E}(x^2x'^2)$$
$$+ \left(-2k^2 + 2kN + 4N - 16k + 12\sum_t \frac{1}{n_t}\right)\mathscr{E}(x^2x'x'')$$
$$+ \left(N^2 - 2kN + k^2 - 4N + 10k - 6\sum_t \frac{1}{n_t}\right)\mathscr{E}(xx'x''x'''). \qquad \ldots\ldots(10)$$

But, remembering that $\sum_j y_j = 0$ and proceeding as in reaching (7) we have,

$$\left.\begin{aligned}\mathscr{E}(x^3x') &= \frac{-\mathscr{E}(x^4)}{(N-1)},\\ \mathscr{E}(x^2x'^2) &= \frac{-\mathscr{E}(x^4)+N\{\mathscr{E}(x^2)\}^2}{(N-1)},\\ \mathscr{E}(x^2x'x'') &= \frac{2\mathscr{E}(x^4)-N\{\mathscr{E}(x^2)\}^2}{(N-1)(N-2)},\\ \mathscr{E}(xx'x''x''') &= \frac{-6\mathscr{E}(x^4)+3N\{\mathscr{E}(x^2)\}^2}{(N-1)(N-2)(N-3)}.\end{aligned}\right\} \qquad \ldots\ldots(11)$$

Also, since $\mathscr{E}(x^2) = \left(\sum_j y_j^2\right)\Big/N$ and $\mathscr{E}(x^4) = \left(\sum_j y_j^4\right)\Big/N$ it follows by (3) and (4) that (11) can be expressed in terms of K_2 and K_4. Making these substitutions into (10) we obtain by straightforward algebra

$$\mathscr{E}(S_1^2) = \frac{K_2^2(N-1)(k^2-1)}{(N+1)} - K_4\left\{\frac{2(k-1)(N-k)}{N(N+1)} + \left(\frac{k^2}{N} - \sum_t \frac{1}{n_t}\right)\right\}. \quad \ldots(12)$$

Whence, by (8),

$$V(S_1^2) = \frac{2(k-1)(N-k)}{(N+1)}\left\{K_2^2 - \frac{K_4}{N}\right\} - K_4\left\{\frac{k^2}{N} - \sum_t \frac{1}{n_t}\right\}. \qquad \ldots\ldots(13)$$

Now
$$E^2 = \frac{S_1}{S(x_{ti}-x_{..})^2} = \frac{S_1}{\sum_j y_j^2} = \frac{S_1}{(N-1)K_2}. \qquad \ldots\ldots(14)$$

Therefore by (8) and (13)

$$\text{Mean } E^2 = \frac{(k-1)}{(N-1)}, \qquad \ldots\ldots(15)$$

$$V(E^2) = \frac{2(k-1)(N-k)}{(N+1)(N-1)^2}\left\{1 - \frac{K_4}{NK_2^2}\right\} - \frac{K_4}{(N-1)^2K_2^2}\left\{\frac{k^2}{N} - \sum_t \frac{1}{n_t}\right\}. \quad \ldots(16)$$

The mean and variance of E^2 for the limited population may usefully be compared with the mean and variance of E^2 derived from (2), which are the appropriate frequency constants when the samples are interpreted as randomly drawn from a normal population. From (2)

$$\text{Mean } E^2 = \frac{(k-1)}{(N-1)}, \qquad \ldots\ldots(17)$$

$$V(E^2) = \frac{2(k-1)(N-k)}{(N+1)(N-1)^2}. \qquad \ldots\ldots(18)$$

(15) agrees exactly with (17), and (16) differs from (18) only in the inclusion of a term K_4/K_2^2. This term will be relatively small if N is large enough and the n_t's not too unequal. (In the particular case where the n_t's are all equal, we shall have k^2/N equal to $\sum_t(1/n_t)$, and (16) will simplify owing to the vanishing of the last term.) In these circumstances no essentially different conclusions will be drawn, whether the samples are regarded as drawn from a limited universe or, as is usual, from an unlimited normal universe.

When N is large in comparison with k, the Beta-function approximation (2) is practically equivalent to the assumption that $(N-1)E^2$ is distributed as χ^2 with $(k-1)$ degrees of freedom. This result has also been found by Geary (1927, p. 106) following a rather different approach to the problem.

An index, which will show whether (2) approximates closely enough the distribution of E^2 in the limited universe, is provided by the ratio, R, of (16) to (18), viz.

$$R = 1 - \frac{K_4}{NK_2^2} - \frac{(N+1)K_4}{2(k-1)(N-k)K_2^2}\left\{\frac{k^2}{N} - \sum_t \frac{1}{n_t}\right\}. \qquad \text{......(19)}$$

The closer this is to unity the better the approximation is likely to be. As an example, Table II gives values of R corresponding to $N = 30$, $k = 3$ and different values of n_1, n_2 and n_3.

TABLE II

n_1	n_2	n_3	R
10	10	10	$1-0{\cdot}033\,K_4/K_2^2$
5	10	15	$1-0{\cdot}014\,K_4/K_2^2$
2	3	25	$1+0{\cdot}131\,K_4/K_2^2$
1	4	25	$1+0{\cdot}251\,K_4/K_2^2$

The table shows how the last term in (19) becomes relatively important when the sample sizes are very disparate, although, up to a point, inequality of sample size has the effect of making R closer to unity. This is due to the fact that $\left(\frac{k^2}{N} - \sum_t \frac{1}{n_t}\right)$ is necessarily non-positive.

Apart from the sample sizes, (19) depends only on K_4/K_2^2. Now it is possible to show that $\left(\sum_j y_j^4\right)\Big/\left(\sum_j y_j\right)^2$ must lie between $1/N$ and $(N^2-3N+3)/N(N-1)$ and hence that K_4/K_2^2 must lie between $-2(N-1)/(N-3)$ and N. Hence limits may be set to the possible values of R. In particular, when all the n_t's are equal we see that R must lie between zero and $1+2(N-1)/N(N-3)$. With N large, therefore, there is in this case no possibility that the variance of E^2 in the limited universe will exceed by much, the normal theory variance. These results are similar to those obtained by me in an investigation into the theory of randomized block experiments, and discussed somewhat fully in a previous number of this journal (Welch, 1937, p. 28).

When R differs sufficiently from unity to make the normal theory approximation inadequate, the question will still remain as to what other method of approximating can be adopted. One such method is to use the true mean and variance of (15) and (16) and fit a Pearson Type I curve with extremities 0 and 1. Alternatively expressions for higher moments may be obtained and used. However, in any attempt to represent the distribution of E^2 in the limited universe by a smooth curve, it must be borne in mind that the distribution is essentially

discrete. Further, it is probable, that it is in just those cases where R differs considerably from unity, that the distribution will tend to be most irregular. Any very elaborate method of fitting a smooth curve may not therefore be justified. With very small samples it will, of course, be possible to evaluate without great difficulty, sufficient of the discrete distribution of E^2, to see where the 5 % level of significance falls. Whether this is worth while, depends on the manner in which the samples are being interpreted.

3. Sampling a more extended population

One argument for the limited universe approach is that it seems to involve a minimum of hypothesis, not assuming anything which is not given directly by the observed sample values. Nevertheless the limited universe is still only a mental concept. It does not have the same concreteness as a population, say, of unemployed workers, from which a certain sample is drawn to provide the basis of a social enquiry. This latter population definitely exists and could be sampled in its entirety if necessary. But a universe generated by shuffling an observed set of samples does not correspond to anything concrete. Only the observed samples are really possible. For example, where a randomized field experiment is carried out, only the treatment actually used on a plot has a corresponding real yield. The other treatments cannot yield figures for that plot at the same time. We can, however, make a mental construct, an hypothesis, as to what they might have been. The hypothesis may be that on every plot the other treatments would have yielded the same as the observed, and this can be tested. The discussion of the previous section will then be relevant. But in cases where there has not even existed the possibility of the observed individuals being classed into groups, other than as they actually are classed, it will not be making any more serious assumptions to interpret the samples in the usual way, as being drawn from an unlimited population, rather than from the constructed limited one. In this section, therefore, we shall consider the appropriate theoretical distribution to which the observed E^2 should be referred, when the k samples are regarded as being drawn randomly and independently from the same infinite population, not, however, necessarily normal.

Use can still be made of the results of the previous section, for all the configurations obtained by shuffling the observed results are still equally likely. We are, in effect, taking the additional step of regarding the aggregate of N as a random sample of N. For the infinite universe therefore we have from (15) and (16)

$$\text{Mean } E^2 = \frac{(k-1)}{(N-1)}, \qquad \ldots\ldots(20)$$

$$V(E^2) = \frac{2(k-1)(N-k)}{(N+1)(N-1)^2}\left\{1-\frac{\alpha_N}{N}\right\} - \frac{\alpha_N}{(N-1)^2}\left\{\frac{k^2}{N}-\sum_t \frac{1}{n_t}\right\}, \qquad \ldots\ldots(21)$$

where α_N is used to denote the expectation, in samples of N, of K_4/K_2^2. (In

formulae (3) and (4), of course, y_j will be replaced by $(x_{ti}-x_{..})$, as the mean $x_{..}$ is now allowed to vary.) Note that in the case where the infinite universe is normal, $\alpha_N = 0$, and (21) becomes (18). In general a corresponding value of R will be obtained by replacing K_4/K_2^2 in (19) by α_N. Since, whatever the sample of N, K_4/K_2^2 is forced to lie within certain limits, so also is α_N. If the population sampled is continuous and of known *shape*, so that α_N is known, then the distribution of E^2 will range continuously from zero to unity with known moments given by (20) and (21). It may then be approximated by the Type I curve

$$p(E^2) = \text{const.} \times (E^2)^{l-1}(1-E^2)^{m-1}, \qquad \ldots\ldots(22)$$

where

$$l = \frac{\mu_1'(\mu_1'-\mu_2')}{(\mu_2'-\mu_1'^2)}; \quad m = \frac{(1-\mu_1')(\mu_1'-\mu_2')}{(\mu_2'-\mu_1'^2)}, \qquad \ldots\ldots(23)$$

μ_1' and μ_2' being the first and second moments of E^2 about zero. More generally α_N will not be known and in that case, an unbiassed estimate of it is provided by K_4/K_2^2. We should then use (16) instead of (21) in (23). If we judge significance from levels calculated in such a way, the levels will change with different aggregates of N and some further investigation is necessary before it can be definitely stated that in the long run we shall be running the stipulated risk (say 5 %) of rejecting the hypothesis of a common source for the k samples, when it is, in fact, true. There is, however, no obvious reason to expect much deviation from this prescribed risk.

4. The χ^2 test for homogeneity of binomial series

This test can be deduced as a particular case of the E^2 test by supposing that x is a variate which equals 1 when the individual has a certain character A, and equals 0 when the individual does not have the character. Let z_t denote the number with character A in the tth sample and let $Z = \sum z_t$. Then

$$S(x_{ti}-x_{..})^2 = Sx_{ti}^2 - \frac{(Sx_{ti})^2}{N} = Z\left(1-\frac{Z}{N}\right)$$

and

$$S(x_{t.}-x_{..})^2 = \sum_t n_t\left(\frac{z_t}{n_t}-\frac{Z}{N}\right)^2,$$

whence

$$E^2 = \frac{\sum_t n_t\left(\frac{z_t}{n_t}-\frac{Z}{N}\right)^2}{Z\left(1-\frac{Z}{N}\right)}.$$

NE^2 is therefore seen to be equal to the measure of dispersion obtained by applying the general χ^2 method of squaring the deviation of each frequency from its estimated expectation, dividing by this expectation and summing over all categories. In another terminology NE^2/k is the Lexis ratio.* In the present discussion we shall denote the above measure of dispersion by D, and the Lexis

* In yet another terminology E^2 is equivalent to the mean square contingency ϕ^2.

ratio by L, and we shall suppose that the sample sizes are equal, although this restriction is not necessary. We then have

$$E^2 = \frac{D}{N} = \frac{L}{n} = \frac{\sum_t (z_t - \bar{z})^2}{k\bar{z}(n-\bar{z})}, \qquad \text{......(24)}$$

where $\bar{z}$ is the mean of z_t.

It is known that when the expectations in the samples are large enough, the distribution of D is well represented by the tabular χ^2-distribution with $f = (k-1)$, but for very small expectations (or at least for very small n) this is known to fail. As recent discussions of the latter case we may instance those of W. G. Cochran (1936) and J. B. S. Haldane (1937) (although Haldane is concerned for the most part with the case where expectations are given *a priori* and are not, as here, estimated from the data). The results of the previous sections throw some light on the conditions under which χ^2 fails and suggest an alternative procedure which may be of value.

In the first place we may note that whether we are considering a system of repetition where the total Z is fixed, or whether we are considering the more extended population where Z also can vary, we have *exactly* from (15)

$$\text{Mean } D = (N \times \text{Mean } E^2) = \frac{(k-1)(N)}{(N-1)}. \qquad \text{......(25)}$$

For the tabular χ^2, the expectation is $(k-1)$, which suggests, perhaps, that we should get better agreement with the tabular χ^2 if we multiply D by $\left(1 - \frac{1}{N}\right)$. However, as the total number of individuals in all the samples will almost certainly be large, this is not the main source of discrepancy. Proceeding to the variance of D, we see, from (16), that its *leading term* is

$$V(D) = N^2 \times V(E^2) = \frac{2(k-1)(N^2)(N-k)}{(N+1)(N-1)^2}. \qquad \text{......(26)}$$

For N large this tends to the tabular χ^2 value $2(k-1)$, only if k is small compared with N, i.e. if the individual samples are large enough. Cases where n is too small occur, for example, where the samples are litters of mice or, as an extreme case, human twins. In such cases, provided, of course, that k is not also very small, it appears likely that to refer the E^2 of (24) to the Beta-function (2), will be a satisfactory alternative procedure. Stated in a slightly different way, this amounts to judging significance by means of Fisher's z test, where

$$z = \tfrac{1}{2}\log_e \left\{ \frac{\sum_t (z_t - \bar{z})^2}{(k-1)} \div \frac{k\bar{z}(n-\bar{z}) - \sum_t (z_t - \bar{z})^2}{(N-k)} \right\} \qquad \text{......(27)}$$

and $f_1 = (k-1)$, $f_2 = (N-k)$.

For example, consider the case $n = 2$, $k = 20$, $N = 40$. Suppose we happen to be sampling a common binomial population whose $p = \frac{1}{4}$, and that sampling is *without restriction* on the total Z. The true distribution of the E^2 of (24) may be

worked out completely. This has been done and the results are presented in Table III. The possible values of E^2 have been grouped and the second column gives the true chance that E^2 should be equal to or greater than the value E_0^2 given in the first column. In the third column are given the corresponding probabilities calculated on the assumption that E^2 is distributed continuously as

$$p(E^2) = \frac{1}{B(\frac{19}{2}, 10)}(E^2)^{\frac{19}{2}-1}(1-E^2)^{10-1}.$$

Bearing in mind the essential discreteness of the true distribution (there are actually only about 100 distinct values of E^2 with probability greater than 0·0001), the approximation would appear to be good. In the fourth column of the table are given approximations to the same probabilities calculated on the assumption that $D = NE^2$ is distributed as χ^2 with $f = 19$. As is expected, the agreement is not now so good at the tails (which are the most important), the variance of the tabular χ^2 being roughly about twice the true variance of D (cf. Cochran, 1936, p. 214).

TABLE III

E_0^2	True $P(E^2 \geqslant E_0^2)$	Beta-function approx. to $P(E^2 \geqslant E_0^2)$	χ^2 approx. to $P(E^2 \geqslant E_0^2)$
0·00	1·0000	1·0000	1·0000
0·25	0·9979	0·9868	0·9529
0·30	0·9830	0·9557	0·8856
0·35	0·9182	0·8892	0·7837
0·40	0·7555	0·7778	0·6573
0·45	0·5902	0·6260	0·5224
0·50	0·3998	0·4540	0·3946
0·55	0·2937	0·2902	0·2843
0·60	0·1962	0·1594	0·1962
0·65	0·0728	0·0728	0·1302
0·70	0·0328	0·0263	0·0834
0·75	0·0125	0·0070	0·0518
1·00	0·0000	0·0000	0·0033

We may conclude that the distribution of E^2 used to test the equality of the means of normal populations, is also useful for judging the significance of the index of dispersion D, when expected frequencies are small due to n being small.

5. Further remarks

In the last section only the leading term in the variance of D was taken into account. It is of theoretical interest to consider the exact expression. Still confining ourselves to the case $n_t = n$, and in the first instance considering the case where the total Z is fixed, we have from (16),

$$V(D) = \frac{2(k-1)(N^2)(N-k)}{(N+1)(N-1)^2}\left\{1 - \frac{K_4}{NK_2^2}\right\}.$$

But since
$$\left(\sum_j y_j^2\right) = S(x_{ti} - x_{..})^2 = Z\left(1 - \frac{Z}{N}\right),$$
and similarly
$$\left(\sum_j y_j^4\right) = Z\left(1 - \frac{Z}{N}\right)\left(1 - \frac{3Z}{N} + \frac{3Z^2}{N^2}\right),$$
we have from (3) and (4)
$$\frac{K_4}{K_2^2} = \frac{\{(N+1) - 6NPQ\}(N-1)}{(N-2)(N-3)PQ},$$
where P has been written for Z/N and $Q = 1 - P$. Therefore
$$V(D) = \frac{2(k-1)(N^2)(N-k)}{(N+1)(N-1)^2}\left\{1 - \frac{(N-1)(N+1-6NPQ)}{N(N-2)(N-3)PQ}\right\}. \quad \ldots\ldots(28)$$

It will be clear from this equation that $V(D)$ will depart considerably from the first term approximation, if either P or Q is very small. The limiting case is $V(D) = 0$ when either P or $Q = 1/N$. In general for N large and P small the multiplier in the curled bracket of (28) is approximately $(1 - 1/NP)$ which can be taken to be unity if NP (i.e. the fixed total Z) is large enough. The maximum of $V(D)$ occurs when $P = \frac{1}{2}$. The multiplier is then $1 + 2(N-1)/N(N-3)$, but this will be close to unity for N large.

Considering next the case where Z is no longer fixed but is allowed to vary in repeated sampling, the variance of D will now be the expectation of (28). This cannot be written down exactly in terms of the population p and q, but will be given approximately by the substitution of p and q for P and Q. The exact expression requires the expectation of $1/PQ$.

6. Summary

The distribution of the correlation ratio E^2 has been considered. In the first instance the mean and variance of E^2 have been derived for the limited universe generated by repartitioning the aggregate of all the samples. From here the step is made to the distribution for an infinite universe, not necessarily normal. Some light is thrown on the range of applicability of 'normal' theory.

The index of dispersion used for testing the homogeneity of binomial series is treated as a particular case. The χ^2 distribution is known to be inapplicable to this index, if the samples are too small. A method of treating this case is suggested.

REFERENCES

Pitman, E. J. G. (1937). *J. Roy. Statist. Soc.* Suppl. **4**, 119.
Geary, R. C. (1927). *Metron.* **7**, 83.
Cochran, W. G. (1936). *Ann. Eugen.* **8**, 207.
Haldane, J. B. S. (1937). *Biometrika*, **29**, 133.
Welch, B. L. (1937). *Biometrika*, **29**, 21.

SOME ASPECTS OF THE PROBLEM OF RANDOMIZATION*

II. AN ILLUSTRATION OF "STUDENT'S" INQUIRY INTO THE EFFECT OF "BALANCING" IN AGRICULTURAL EXPERIMENTS

By E. S. PEARSON

1. Introductory

In § 4 of his last paper on "Comparison between balanced and random arrangements of field plots" ("Student", 1937), the late Mr W. S. Gosset set before his readers one of those simple yet fruitful ideas which have been so characteristic of his contributions to statistics during a period of thirty years. The section in question was entitled "The effect of 'balancing' on the 'validity' of conclusions". The matter dealt with in this and the preceding sections may fairly be said to bristle with topics for controversy. Nevertheless the fact that "Student", who had an intense dislike of controversy, felt at last impelled to set down on paper his views hereon, is evidence of the strength of his conviction that some protest must be made against the claim, so often repeated, that without randomization the results of experiments are invalid.

In the last few months copies of letters exchanged between "Student" and his agricultural correspondents scattered over the world have come into my possession; they show well what an exceedingly helpful correspondent he was and at the same time make clear that he gained much himself from this long-range exchange of ideas, as he would himself have been the first to admit. It was therefore interesting to discover from a letter to a friend in Australia, written on 7 March 1937, the actual date on which "Student" saw in a flash the consequences, described in the section of his paper to which I have referred, of balancing treatments in a randomized block type of experiment. The genesis of his idea seems to have lain in a re-examination of some experimental analyses of uniformity trial data which Mr A. W. Hudson of Massey College, New Zealand, had sent to him as long ago as October 1933. "This" [a study of Hudson's data], he wrote in the letter to Australia, "put me on to a great truth which should, of course, have been obvious if one had only thought about it." And later: "Sorry to bore you with all this, but I only got hold of it yesterday!"

In the present paper I shall attempt to investigate a little more fully, and in as objective a manner as possible, the central idea that "Student" had in mind. In his paper he applied it to the case of the randomized block and the half drill strip lay-out. I shall deal here with the former problem, and after setting out

* For an earlier paper under the same general title, see E. S. Pearson (1937).

rather more fully than he did the algebraic background of his result, shall illustrate its meaning with the help of diagrams on some of the data used by Mr Hudson in his Appendix ("Student", 1937, pp. 376–9).

2. The randomization set of treatment patterns

Suppose that in an agricultural experiment designed to compare k "treatments", the experimental area is laid out in n blocks each containing k plots. The yield from the jth plot on the ith block may be denoted by x_{ij},

$$(i = 1, \ldots, n; \quad j = 1, \ldots, k),$$

while the yield from the plot in this block receiving the rth treatment will be denoted by $x_{i(r)}$ $(r = 1, \ldots, k)$. Thus the subscript j indicates the position of the plot while r indicates the treatment it receives. Were it desired to indicate that the (ij)th plot receives the rth treatment we could write $x_{ij(r)}$. The analysis of variance procedure carried out to test whether there are significant treatment differences will then consist in calculating the sums of squares shown in the following table:

TABLE I

		Degrees of freedom
Treatment	$S_1^* = \sum_r n(x_{.(r)} - x_{.(.)})^2$	$k-1$
Error	$S_2^* = \sum_{i,r} (x_{i(r)} - x_{.(r)} - x_{i(.)} + x_{.(.)})^2$	$(n-1)(k-1)$
Total	$S_3^* = \sum_{i,r} (x_{i(r)} - x_{i(.)})^2 = \sum_{i,j} (x_{ij} - x_{i(.)})^2$	$n(k-1)$

Here $x_{i(.)}$, $x_{.(r)}$ and $x_{.(.)}$ stand as usual for the block means, the treatment means and grand mean, respectively. If now there are no treatment differences whatsoever, and $x_{i(r)}$ may be considered as made up of a block term plus a normal random residual, say

$$x_{i(r)} = \beta_i + v_{i(r)}, \qquad \ldots\ldots(1)$$

then the probability distribution of

$$u = \frac{S_1}{k-1} \Big/ \left\{ \frac{S_2}{(n-1)(k-1)} \right\} \qquad \ldots\ldots(2)$$

is of well-known form,† which may be termed the "normal theory" distribution of the ratio of two independent estimates of a common variance.

Because experimentalists have been doubtful of the justification of supposing that the $v_{i(r)}$ of equation (1) would in practice, when there are no treatment differences, be independent normal residuals, it has been customary to emphasize the importance of randomly assigning the treatments to the k plots within each

* In what follows S_1, S_2 and S_3 will be used to denote these sums of squares *only in the case where there are no real treatment differences*, e.g. when the analysis is applied to uniformity trial data.

† For purposes of this discussion it is simpler to deal with the quantity u, rather than with $z = \frac{1}{2} \log_e u$.

block. It will be seen that there are $(k!)^{n-1}$ possible partitions of the nk plots into k undifferentiated groups of n, such that each group contains a plot in every block; these may be termed the randomization set of treatment partitions or patterns. When a pattern has been selected there will still be $k!$ ways of laying down k specific treatments; the first treatment, say A_1, may be placed on any one of the k groups of plots, A_2 on any one of the remaining $k-1$ groups, and so on. There are therefore in all $(k!)^n$ possible arrangements* of k treatments. One of these arrangements will have been selected for the experiment. In order to test the hypothesis that the treatments are equivalent, the value of u of equation (2), resulting from this experiment, may then be referred to the set of $(k!)^{n-1}$ values which would be obtained if all the treatment patterns of the randomization set were applied to the observed plot yields x_{ij}. This set of values of u constitutes what may be termed the distribution of u under randomization. As Eden & Yates (1933) suggested experimentally and Welch (1937) and Pitman (1937 b) have shown by more extensive investigation, if there are no real treatment differences the distribution of u under randomization is unlikely to differ seriously from the normal theory form. The total sum of squares S_3 of Table I will be the same in every case, but the apportionment of the total into the parts S_1 and S_2 will vary according to the pattern used.

As an illustration of the points under discussion, I have shown in Table II two of the treatment patterns (or arrangements)† applied by Hudson‡ to Mercer and Hall's uniformity trial data for mangolds (see "Student" (1937), Appendix, Table I, 2nd row). There are four hypothetical treatments a_1, a_2, a_3 and a_4 arranged in 10 blocks; the 40 plot yields given in pounds are shown below the treatment letters. The first arrangement was obtained by Hudson randomly, the second is a balanced arrangement; both patterns associated with the arrangements belong of course to the set of $(4!)^9$ possible patterns of the randomization set. Comparable with Table I, we have the analyses of variance shown in Table III.

It is seen that S_1 is considerably smaller for the balanced than for the random arrangement; consequently u is also smaller in the former case. Neither value of u is however significant. "Student" emphasized the fact that out of the randomization set, balanced arrangements would on the whole be associated with the smaller values of S_1 and consequently larger values of S_2; in other words balancing

* The distinction between the number of treatment patterns and the total number of arrangements is of no importance when treatment differences do not exist. Its meaning when they are present will be discussed more fully in § 3 below.

† These are "patterns" if we think of a_1, a_2, ..., as mere indices of the plot groups; they are "arrangements" if we associate them with specific treatments, e.g. $a_1 = A_2 =$ sulphate of ammonia, $a_2 = A_4 =$ nitrate of soda, etc. Clearly there would be $4! = 24$ ways of associating the indices a with the real treatments A. In so far as we are assigning hypothetical treatments to uniformity trial data the distinction is of no importance, and following "Student's" terminology we shall speak in this section of "arrangements".

‡ I should like to thank Mr Hudson very warmly for looking out his original working sheets and forwarding them to me from New Zealand. I am also glad of the opportunity of making further use of computations into which he must have put an immense amount of labour a few years ago.

TABLE II

Hudson's allocation of treatments, I, 2

Random arrangement (East side on the left, North at the foot)

a_1 1401	a_3 1403	a_1 1339	a_1 1373
a_3 1312	a_4 1335	a_2 1334	a_4 1325
a_2 1337	a_2 1325	a_3 1310	a_3 1264
a_4 1397	a_1 1332	a_4 1304	a_2 1295
a_4 1380	a_4 1314	a_2 1309	a_2 1387
a_1 1373	a_3 1260	a_4 1314	a_1 1375
a_3 1388	a_2 1272	a_1 1222	a_4 1272
a_2 1268	a_1 1290	a_3 1268	a_3 1333
a_1 1310	a_3 1293	a_4 1274	a_2 1321
a_4 1276	a_2 1239	a_1 1215	a_3 1175

East

North

	Plot means	Deviation from grand mean
a_1	1323·0	+10·2
a_2	1308·7	− 4·1
a_3	1300·6	−12·2
a_4	1319·1	+ 6·3
Grand mean	1312·85	

Balanced arrangement (West side on the right, North at the foot)

a_1 1401	a_4 1403	a_3 1339	a_2 1373
a_2 1312	a_3 1335	a_4 1334	a_1 1325
a_3 1337	a_2 1325	a_1 1310	a_4 1264
a_4 1397	a_1 1332	a_2 1304	a_3 1295
a_1 1380	a_4 1314	a_3 1309	a_2 1387
a_2 1373	a_3 1260	a_4 1314	a_1 1375
a_3 1388	a_2 1272	a_1 1222	a_4 1272
a_4 1268	a_1 1290	a_2 1268	a_3 1333
a_1 1310	a_4 1293	a_3 1274	a_2 1321
a_2 1276	a_3 1239	a_4 1215	a_1 1175

West

North

	Plot means	Deviation from grand mean
a_1	1312·0	−0·8
a_2	1321·1	+8·3
a_3	1310·9	−1 9
a_4	1307·4	−5·4
Grand mean	1312·85	

TABLE III

Significance levels for u: 5%, 2·96; 1%, 4·60.

	f		Sum of squares		Mean squares	
			Random	Balanced	Random	Balanced
Treatment	3	S_1	3093·7	1022·9	1031·2	341·0
Error	27	S_2	54113·3	56184·1	2004·2	2080·9
Total	30	S_3	57207·0	57207·0	u=0·514	u=0·164

would tend to reduce the bias in the treatment means, $x_{.(r)}$, due to soil heterogeneity. The result of Hudson's investigation bore out this contention; out of fifteen experiments the balanced lay-outs gave a smaller S_1 than the random on twelve occasions, the reduction being very considerable in some cases. As a consequence, when there are no real treatment differences, the distribution of the ratio of estimates of variance, u, is unlikely for balanced arrangements to follow even approximately the normal theory form. There is certainly no harm in this result when treatment differences do not exist, for nothing is gained by believing once in twenty times that a difference exists when it does not. The real question, however, is what effect will the tendency of obtaining larger values of S_2 among balanced arrangements have upon the efficiency of the test in detecting the presence of real treatment differences when they exist? It was on this point that "Student's" work has thrown new light.

3. "Student's" method of comparing the efficiency of balanced and random arrangements

In dealing with the position when real treatment differences exist, it is necessary to extend somewhat the ideas and notation discussed in the preceding section. It will be noticed that in laying down the experiment an opportunity for choice has occurred at two stages:

Stage 1. It has been necessary to select a particular treatment pattern out of the randomization set of $(k!)^{n-1}$ patterns. Two such patterns were shown in Table II, the particular groups of plots to be associated with the same treatment being indexed by the letters $a_1, a_2, \ldots, a_k$. These may be conveniently described as dummy treatments.

Stage 2. It is further necessary to decide how to associate the k specific treatments under investigation, say $A_1, A_2, \ldots, A_k$ with the dummy treatments $a_1, a_2, \ldots, a_k$. There will be $k!$ ways of doing this. If there are no real treatment differences, as when applying a hypothetical treatment pattern to uniformity trial data, it is immaterial which of the $k!$ alternative ways of associating the a's and A's we make, but in actual practice when laying out an experiment this second choice must be made, and presumably it will be quite randomly made.*

"Student's" approach to the problem was as usual very simple; it consisted essentially in two steps. In the first place he suggested that the position could be explored by regarding the plot yields as represented by what amounts to the following symbolic equation:

$$x_{i(rs)} = m_{i(r)} + \delta_s. \qquad \ldots\ldots(3)$$

Here $x_{i(rs)}$ represents the yield from that plot in the ith block which at stage 1, in choosing the pattern, was assigned dummy treatment a_r and at stage 2

* The experimentalist choosing a random arrangement will no doubt often combine the two stages and make a single choice. If, however, at the first stage he selects some pattern, say, from a printed series, the second choice has still to be made.

received the real treatment A_s. It is built up of two additive parts; the first part, $m_{i(r)}$, is associated with the plot in the ith block to which a_r has been assigned and would be the same whatever real treatment were applied; the second part, δ_s, would be the same for treatment A_s on all plots. The two subscripts r and s have been introduced to indicate that at stage 2 there are $k!$ ways of associating $\delta_1, \ldots, \delta_s, \ldots, \delta_k$ with the plots indexed by the dummy treatments $a_1, \ldots, a_r, \ldots, a_k$ of a particular treatment pattern. It is seen from equation (3) that the term $m_{i(r)}$ will vary from plot to plot in exactly the same manner as would be found in a uniformity trial. If we suppose $\sum_s (\delta_s) = 0$, then $m_{i(r)}$ is the average of the yields we should expect to find if it were possible to apply all k treatments in turn to a plot under the same conditions.

Clearly an assumption is involved in equation (3), since it is supposed that the contribution δ_s is the same on all plots treated with A_s, whereas in fact there might well be some interaction between the treatment and the soil characteristics of a plot. Again since only one treatment will in fact be applied to any single plot, all combinations of $m_{i(r)}$ and δ_s, except one, will be hypothetical. It must be remembered, however, that any probability statements whatsoever that can be made regarding the test criterion u must depend on the construction of some conceptual model of this kind, and "Student's" set-up needs no special pleading.

If now we write $x_{(rs)}$ for the mean yield of plots receiving A_s, when this treatment is associated with the dummy treatment a_r of a particular randomization pattern; $m_{.(r)}$ as the mean value of $m_{i(r)}$ on these plots; and other mean values as in § 2 above, we shall have

$$x_{.(rs)} = m_{.(r)} + \delta_s, \quad x_{i(.)} = m_{i(.)}, \quad x_{.(.)} = m_{.(.)}. \qquad \ldots\ldots(4)$$

The analysis of Table I applied to the x's will give:

TABLE IV

		Degrees of freedom
Treatment	$S_1' = \sum_r n(m_{.(r)} - m_{.(.)} + \delta_s)^2$	$k-1$
Error	$S_2 = \sum_{i,r} (m_{i(r)} - m_{.(r)} - m_{i(.)} + m_{.(.)})^2$	$(n-1)(k-1)$
Total	$S_3' = \sum_{i,r} (m_{i(r)} - m_{i(.)} + \delta_s)^2$	$n(k-1)$

Here the error term, S_2, depends only on the m's and its behaviour under randomization at stage 1 we have already discussed in the preceding section. The treatment term S_1'* breaks up into three parts, the first of which is

$$S_1 = \sum_r n\,(m_{.(r)} - m_{.(.)})^2$$

* In this notation the convention referred to in connexion with Table I is being retained, namely S_1 and S_2 relate to sums of squares of terms containing no real treatment differences.

also depending only on the m's. These two terms will, on "Student's" hypothesis, remain the same whatever the combination at stage 2.

The test criterion u may now be expressed as

$$u = \frac{(n-1)\,S_1'}{S_2} = \frac{(n-1)\,S_1 + n(n-1)\left\{\sum_s(\delta_s^2) + 2\sum_r(m_{\cdot(r)} - m_{\cdot(\cdot)})\,\delta_s\right\}}{S_2} \qquad \ldots\ldots(5)$$

The second step in "Student's" approach was as follows; by selecting a balanced pattern, the random element has been removed at stage 1, but a random choice remains at stage 2. Thus starting from a basic set of m's, and a given treatment pattern, there will still be $k!$ possible values of u depending on the way in which the elements in the product-sum in the numerator on the right-hand side of equation (5) are associated. Any one of the values of u will be equally likely to arise, on his hypothesis, since the treatment terms δ_s will bear no relation to the terms $m_{\cdot(r)} - m_{\cdot(\cdot)}$ representing bias due to soil heterogeneity.

In §4 of his paper, "Student" has used the following terminology:

$$\left.\begin{aligned} &\text{(i)}\ \sigma_e^2 = \frac{\sum_r(m_{\cdot(r)} - m_{\cdot(\cdot)})^2}{k-1} = \frac{S_1}{n(k-1)}, \quad \text{(ii)}\ \sigma_c^2 = \frac{S_2}{n(n-1)(k-1)}, \\ &\text{(iii)}\ \sigma_T^2 = \frac{\sum_s(\delta_s^2)}{k-1}, \end{aligned}\right\} \qquad \ldots\ldots(6)$$

and has spoken of these as (i) the actual variance of error, (ii) the calculated variance of error, and (iii) the real variance due to treatment. In this notation we may write

$$u = \frac{\sigma_e^2}{\sigma_c^2} + \frac{\sigma_T^2}{\sigma_c^2} + \frac{2\sigma_e\sigma_T r_{Te}}{\sigma_c^2}, \qquad \ldots\ldots(7)$$

where r_{Te} is the coefficient of correlation between $m_{\cdot(r)} - m_{(\cdot)}$ and δ_s. It should be noted that

$$n(k-1)\,\sigma_e^2 + n(n-1)(k-1)\,\sigma_c^2 = \sum_{t,r}(m_{t(r)} - m_{t(\cdot)})^2 = S_3, \qquad \ldots\ldots(8)$$

where, for all the randomization sets of a given series of m's, S_3 is constant.

The existence of treatment differences will be detected when u falls beyond the particular significance level chosen. To show the effect of balancing on the efficiency of the test, "Student" took the case $k = 4$, $n = 6$ and supposed it possible to pick out from the $(4!)^5$ possible arrangements of treatments three which, when applied to the basic m's, made

$$\left.\begin{aligned} &(a)\ \sigma_e^2 = \sigma_c^2 = S_3/n^2(k-1) = \sigma^2\ \text{(say)}, \\ &(b)\ \sigma_e^2 = 0{\cdot}5\,\sigma^2,\ \sigma_c^2 = 1{\cdot}1\,\sigma^2, \\ &(c)\ \sigma_e^2 = 1{\cdot}5\,\sigma^2,\ \sigma_c^2 = 0{\cdot}9\,\sigma^2. \end{aligned}\right\} \qquad \ldots\ldots(9)$$

It will be noticed that (*b*) and (*c*) as well as (*a*) satisfy equation (8) which, for $k = 4$, $n = 6$, becomes

$$18\sigma_c^2 + 90\sigma_c^2 = S_3 = 108\sigma^2. \qquad \text{......(10)}$$

The variation in the u of equation (7) will depend on the variation in r_{Te} under randomization at stage 2. The distribution of this coefficient is of the type which we should find if we took two series of numbers say u_1, u_2, u_3, u_4 and v_1, v_2, v_3, v_4, for which $\sum_i (u_i) = 0 = \sum_j (v_j)$, and calculated the 24 possible correlations

$$\sum_{i,j}(u_i v_j) / \sqrt{\sum_i (u_i^2) \sum_j (v_j^2)}$$

arising from the 24 possible pairings of the u's and v's. "Student" supposed*

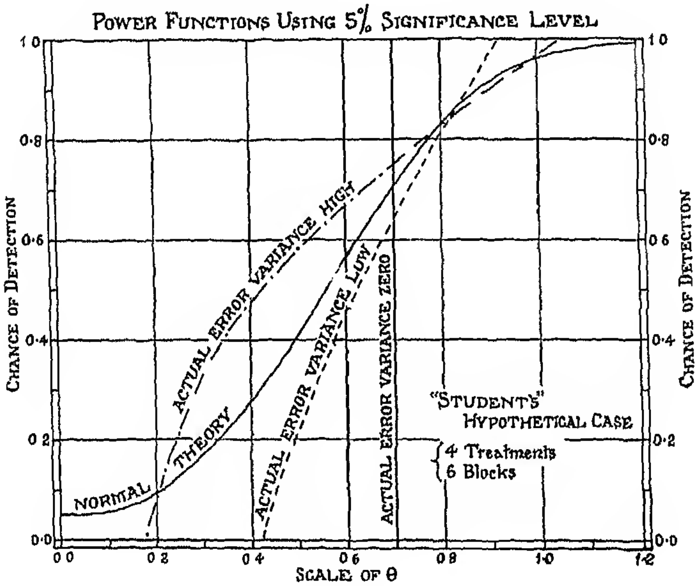


Fig. 1.

that the distribution would be that of a correlation coefficient in a sample of four from a normal bivariate population with correlation $\rho = 0$, that is to say that r_{Te} would be equally likely to assume any value between -1 and $+1$. On this assumption he was able to calculate readily the chance that the u defined in equation (7) would fall beyond the 5% significance level (in this instance at 3·287). The result of these calculations for the cases (*a*), (*b*) and (*c*) are shown in the table on p. 373 of his paper. These chances of detection are shown, for the extreme cases (*b*) and (*c*) in my Fig. 1; in this presentation of the results two points should be noted:

* While this is not *strictly true*, *Pitman (1937a)* has shown how rapidly the distribution of r under randomization approaches that of normal theory as the number of elements is increased. See the case he illustrates with $k = 5$.

(1) I have taken as the measure of real treatment differences

$$\theta = \frac{\sqrt{\frac{1}{k}\Sigma(\delta_s^2)}}{\sqrt{n}\,\sigma} = \frac{\sigma_T}{\sigma}\sqrt{\frac{k-1}{nk}}, \qquad \ldots\ldots(11)$$

which is the ratio of the standard deviation of treatment differences, to the estimate of the standard error per plot (the standard deviation of the $v_{i(r)}$'s of equation (1)) which we should obtain when $\sigma_e = \sigma_c$.

(2) I have described the chance of detection of treatment differences using a given test of significance as the "power" of the test, and the curves as "power functions". This I have done to conform with the terminology used by J. Neyman & E. S. Pearson (1936), in discussing this aspect of tests of statistical hypotheses from the general theoretical view-point. The third curve added to the diagram and described as that of normal theory, will be referred to in § 5 below.

Since "Student" had pointed out that balancing was likely on the average to give lower values to $\sigma_e^2 = S_1/n\,(k-1)$ than a random assignment of treatments, his conclusions may be simply illustrated on this diagram. The curves, which represent the chance of detecting treatment differences plotted against θ, will rise more and more steeply the smaller is S_1. Should $S_1 = 0$, the curve becomes in the limit a vertical line rising from the point* $\theta = \{v_{0\cdot05}(k-1)/k(n-1)\}^{\frac{1}{2}}$, which in the present example is 0·702, $u_{0\cdot05}$ being the 5 % significance level. A steep curve is associated with a zero chance of detecting small treatment differences, but as θ increases it will lead to a chance approaching unity more rapidly than for a curve of lesser slope. The two dotted curves in the diagram cross at about $\theta = 0\cdot82$. The properties of these steeper curves are therefore likely to be associated with balanced lay-outs. How far these properties are advantageous or otherwise, will be discussed later.

4. FURTHER ILLUSTRATIONS USING HUDSON'S DATA

The practical implications of "Student's" argument will clearly depend on how far a difference between σ_e^2 and σ_c^2 of the magnitude indicated in equation (9), case (b), is likely to follow from balancing the treatments in the blocks. To investigate this point it seemed desirable to apply his method to certain of the treatment arrangements used by Hudson. The process which I have followed consists, in effect, of building up hypothetical trials by adding treatment differences, δ_s, as in equation (3), to the uniformity trial plot yields used by Hudson, which will now be the $m_{i(r)}$'s in the notation of § 3. The result may be first illustrated on the example set out in Tables II and III above, for which $k = 4$, $n = 10$.

Instead of using the expression for u in the form (7), let us return to the form (5). For any given set of four values $m_{.(r)} - m_{.(.)}$ $(r = 1, \ldots, 4)$, such as those for the random arrangement of Table II, namely,

$$+10\cdot2, \quad -4\cdot1, \quad -12\cdot2, \quad +6\cdot3, \qquad \ldots\ldots(12)$$

* This follows from setting $\sigma_e = 0$ in equations (7) and (8) and then using (11).

and a set of four real treatment differences δ_s, such as

$$+50, \quad 0, \quad -20, \quad -30, \qquad \ldots\ldots(13)$$

there will be $4! = 24$ different values of the numerator on the right-hand side of equation (5). These correspond to the 24 ways in which the series (12) and (13) may be paired to form $\Sigma(m._{(r)} - m._{(\cdot)})\,\delta_s$. Any one of these may be regarded as equally likely to arise in practice, since the assignment of particular treatments to the plots marked a_1, a_2, etc., in Table II will be entirely random. Since for the series of treatment errors (12), S_2 is given in Table III as 54113·3, it is easy to calculate the resulting 24 values of u. These values will vary about a mean of

$$\bar{u}(\sigma_t^2) = \frac{(n-1)\{S_1 + n\Sigma(\delta_s^2)\}}{S_2} \qquad \ldots\ldots(14)$$

$$= 0{\cdot}5143 + 0{\cdot}006650\sigma_t^2,$$

where $$\sigma_t^2 = \tfrac{1}{4}\Sigma(\delta_s^2).\text{*} \qquad \ldots\ldots(15)$$

This straight line is shown in the upper portion of Fig. 2, in a diagram whose coordinate axes are u and σ_t^2. For a given σ_t^2 and S_1 (or $S_2 = S_3 - S_1$), the variation of the 24 values of u about their mean is proportional to σ_t. Retaining the same *relative* magnitude and sign for the δ_s, as given in (13), but using an adjustable scaling factor, it was easy to calculate the 24 values of u appropriate for various values of σ_t^2. These are shown as arrays of spots in the diagram; the 5 % and 1 % significance levels for u have also been drawn.

The same process, using the same set of values for σ_t^2 was applied to the balanced lay-out of Table II. We now start with treatment errors

$$-0{\cdot}8, \quad +8{\cdot}3, \quad -1{\cdot}9, \quad -5{\cdot}4, \qquad \ldots\ldots(16)$$

and $S_2 = 56184{\cdot}1$. The mean of the 24 values of u is now

$$\bar{u}(\sigma_t^2) = 0{\cdot}1640 + 0{\cdot}006412\sigma_t^2. \qquad \ldots\ldots(17)$$

The situation is readily understood from a comparison of the two charts. As $\sigma_t^2 \to 0$ the 24 possible u-values close in towards one another, and when $\sigma_t^2 = 0$ we have $u = 0{\cdot}514$ for the random and $0{\cdot}164$ for the balanced arrangement. Neither of these values are significant. As σ_t^2 increases some of the u's begin to fall beyond the 5 % level; this occurs sooner in the random than in the balanced case, partly owing to $\bar{u}(\sigma_t^2)$ being larger and partly owing to the greater spread of the 24 u's which depends on S_1. When, however, $\bar{u}(\sigma_t^2)$ for the balanced case falls beyond the 5 % (or 1 %) level, the smaller spread in the u's, resulting from the smaller S_1, is advantageous, and the chance of detecting the existence of real treatment differences is greater than for the random case. A count of the number of values of u falling beyond the two significance levels for different values of σ_t, leads to the results shown in Table V, which illustrates the crossing over of the power curves, previously seen in Fig. 1.

* Note that this quantity σ_t differs by a constant factor from "Students" σ_T defied in equation (6) (iii).

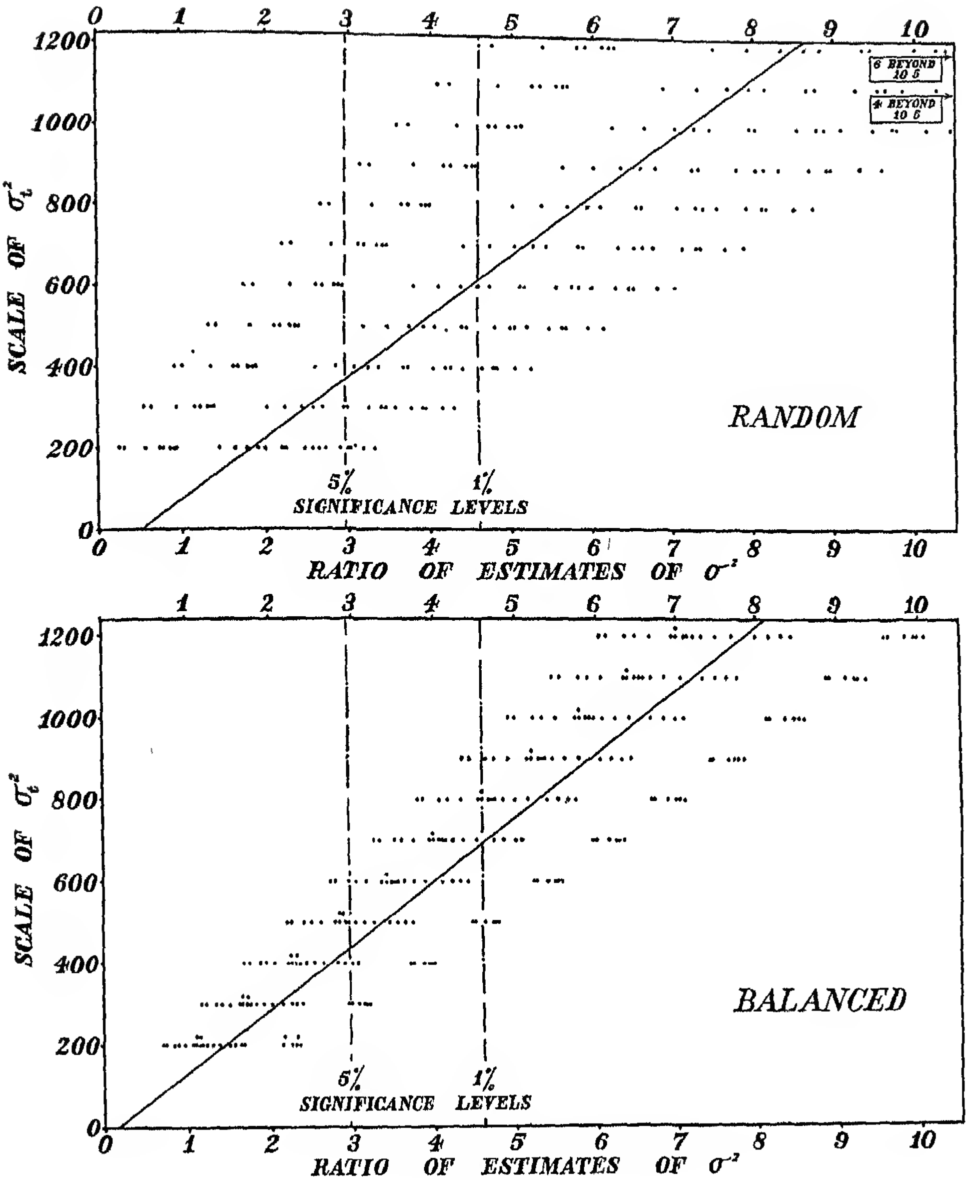


Fig. 2.

The arrangement of treatments shown in Table II are of course only two of the $(4!)^9$ alternatives of the randomization set. Each of these will have its regression line

$$\bar{u}(\sigma_t^2) = u_0 + b\sigma_t^2, \qquad \ldots\ldots(18)$$

and since as S_1 increases, u_0 and b increase, these lines will not cut. Approximately

TABLE V

σ_t^2	Frequency of u above			
	5 % significance level		1 % significance level	
	Random	Balanced	Random	Balanced
100	0	—	—	—
200	4	0	—	—
300	10	5	0	—
400	14	8	5	0
500	16	13	9	4
600	16	22	13	6
700	21	24	15	10
800	22	—	16	18
900	24	—	16	22
1000	—	—	21	24
1100	—	—	22	—
1200	—	—	24	—

5 % of the values of u_0 will fall beyond the level $u_{0\cdot05}=2\cdot96$, and 1 % beyond $u_{0\cdot01}=4\cdot60$. Further, the spread of the 24 u's in the arrays will depend on S_1 and σ_t.

Although this simple method of presenting the situation was not mentioned in "Student's" paper as published, it was outlined by him in correspondence on the subject a few months before his death.

The process of calculating the $k!$ possible product sums of the differences $m._{(r)}-m._{(\cdot)}$ and δ_s becomes very lengthy when $k>4$. Luckily in this connexion Dr L. J. Comrie and Mr G. B. Hey came to my assistance with a scheme that could be easily worked with the Hollerith Calculating Machine. It was therefore possible to carry out the same procedure on a number of Mr Hudson's random and balanced lay-outs involving six treatments and, therefore, $6! = 720$ possible product sums. The method by which the data for Figs. 3–6 were obtained is described in an Appendix. All that need be stated here is that a basic set of hypothetical treatment differences $d_s(s = 1, \ldots, 6)$, was first selected and the product sums $\Sigma(m._{(r)}-m._{(\cdot)})\,d_s$ determined. Then writing $\delta_s = qd_s$, it was easy to adjust these product sums to correspond with any desired value of σ_t, since $q = \sigma_t/\sigma_d$, where

$$\sigma_d^2 = \frac{1}{k}\Sigma(d_s^2), \quad \sigma_t^2 = \frac{1}{k}\Sigma(\delta_s^2). \qquad \ldots\ldots(19)$$

Table VI shows four series of values of d_s which were introduced as described below. Table VII gives the essential particulars of Hudson's cases used. In each case S_1 for the balanced arrangement is less than for the corresponding random arrangement. This will certainly not always be the case in practice; my purpose is,

TABLE VI

Values of d_s used in experiments

	Series 1	Series 2	Series 3	Series 4
d_1	+6	+5	+5	+5
d_2	0	+1	+2	+4
d_3	−1	0	0	−1
d_4	−1	−1	−1	−2
d_5	−2	−1	−2	−3
d_6	−2	−4	−4	−3
σ_d	2·7689	2·7080	2·8867	3·2660

TABLE VII

Data from Hudson's arrangements of six treatments in randomized blocks

	Values of $m_{\cdot(r)} - m_{\cdot(\cdot)}$								
Exp.*...	Table II, No. 4		Table III, No. 2		Table III, No. 4		Table III, No. 6		
	Random	Balanced	Random (B)	Balanced	Random (A)	Balanced	Random (A)	Random (B)	Balanced
$r=1$	+68·2	+11·4	+1·1	−1·4	+8·8	+3·0	−12·1	− 5·9	− 3·3
2	+40·5	+ 5·4	+1·8	+1·3	+0·9	+3·3	− 7·2	+39·0	−10·8
3	−41·6	−17·0	−3·1	+0·9	−5·4	−2·0	+41·0	+10·8	+ 5·6
4	− 6·3	+ 2·1	+1·2	+0·7	−1·4	−3 3	+35·8	−16·9	− 0·6
5	−73·3	− 5·4	+1·1	0·0	−2·8	0·0	−19 0	−12·5	+ 1·1
6	+12 5	+ 3·6	−2·1	−1·5	−0·1	−1·0	−38 7	−14·6	+ 7·8
S_1	54379·2	1984·3	330·2	118·1	927·8	283·9	20070·6	9316·2	905·3
S_2	33650·1	86045·0	2676·8	2889 0	8023·6	8667·5	66375·4	77129·8	85540·7
S_3	88029·3		3007 1		8951·4		86446·0		
n	4		16		8		4		

* The table references are those in Hudson's Appendix ("Student", 1937, pp. 377–9). Table II deals with a uniformity trial of sugar beet (Immer, 1932), and Table III with a uniformity trial for potatoes (Kalamkar, 1932). Note that n, as in the text, indicates the number of blocks; $k=6$ throughout.

however, in the first place to investigate the nature of the differences in the power function curves which result from differences actually met with in S_1 by Hudson within the randomization set.

In drawing the diagrams the chance of detection of treatment differences, or the power of the test, might be plotted against σ_t^2 or σ_t. To bring the diagrams into standardized form and to enable comparison to be made with the normal theory curves, described below, it would be desirable to take as abscissa the ratio of σ_t to the true standard error per plot. The latter is however of course unknown, and all that is possible is to use some estimate of its value. For this I have taken $\sigma' = S_2/(n-1)(k-1)$, using S_2 from the random arrangement, so that θ in the diagram is given by

$$\theta = \sigma_t/\sigma'. \qquad \text{......(20)}$$

It might have been better to take σ' as $S_3/n(k-1)$, as when discussing "Student's" hypothetical case on pp. 165–7 above, but this had not struck me until after the diagrams were drawn. The main point, however, is that the same value of σ' must be used in comparing the efficiency of the random and balanced arrangements.

The four cases considered may now be described in detail.

Fig. 3 (Hudson, II, 4). Series 2 of d_s values from Table VI were used. The curves show the chance of detection of treatment differences for random and balanced arrangements when the hypothesis $\sigma_t = 0$, is rejected if (i) $u > u_{0\cdot 05} = 2\cdot 901$, (ii) $u > u_{0\cdot 01} = 4\cdot 556$. Since for the random arrangement in the original uniformity trial

$$u_0 = \frac{S_1}{5} \Big/ \frac{S_2}{15} = 4\cdot 85,$$

and therefore lies beyond the 1 % limit, the form of the power curve is peculiar. Using the 5 % level of significance, we are certain to detect differences between $\theta = 0$ and $\theta = 0\cdot 24$; larger differences will sometimes be overlooked though the chance is never less than 9 to 1 against this. When $\theta \geqslant 1\cdot 66$ we shall again be certain of detecting differences. For the balanced arrangement, using the 5 % level, we shall detect no differences until $\theta = 1\cdot 07$. From this point the curve rises rapidly and when $\theta \geqslant 1\cdot 43$ treatment differences will be certainly detected. It will be noticed that "certainty" is secured for the balanced at a slightly lower value than for the random arrangement; this is what "Student" expected, but he had not perhaps realized the peculiar nature of the power curve for lower values of θ in this case where u is significant for $\sigma_t = 0$.

The curves shown result, of course, from the particular series of basic d_s values used, namely the series 2 of Table VI. To examine what change in the curves would result if the distribution of real treatment differences were changed, similar calculations were made, using series 1. This series has a single exceptional high value, d_1, the other five values being close together. Table VIII shows a comparison of the chances of detection for corresponding values of $\theta = \sigma_t/\sigma'$; there is seen to be relatively little difference between the figures in the corresponding columns

TABLE VIII

Random					Balanced				
	Chance of detection using					Chance of detection using			
	5% level		1% level			5% level		1% level	
d-series →	1	2	1	2	d-series →	1	2	1	2
θ=0·00	1·000	1·000	1·000	1 000	θ=1·10	0·000	0·039		
0·05	1·000	1·000	0·886	0·913	1·15	0·163	0·186		
0·10	1·000	1·000	0·744	0·740	1·20	0·467	0·414		
0·15	1·000	1 000	0·694	0·674	1·25	0·672	0·617		
0·25	1 000	0·997	0·684	0·665	1·30	0·799	0·767		
0·40	0·940	0·936	0·694	0·675	1·35	0·833	0·878	0·000	0·000
0·60	0 906	0·914	0·744	0·739	1·40	0·933	0·967	0·000	0·024
0·80	0·906	0·918	0·780	0·806	1·45	1·000	1·000	0·100	0·125
1·00	0·933	0·936	0·822	0·858	1·50			0 333	0·315
1·20	0·967	0·965	0 890	0·914	1·55			0·621	0·536
1·40	1·000	0 983	0·951	0·943	1·60			0 761	0·732
1·60	1 000	0·998	0·997	0·978	1·65			0·833	0·850
1·80	1·000	1·000	1·000	0·994	1·70			0·869	0·939

headed "1" and "2". In other words for a given value of S_1, the chance of detection of treatment differences depends mainly on the standard deviation of the δ_s and very little on the form in which they are distributed.

Fig. 4 (Hudson, III, 2). Series 3 of the d_s values from Table VI was used; it differs only very slightly from series 2. The power curves are shown for a random and balanced arrangement using the 1 % significance level for u, which for $k=6$, $n=16$ is $u_{0\cdot01}=3\cdot271$. Owing to the large number of replications, the curves rise steeply; the cross-over effect is again present.

Fig. 5 (Hudson, III, 4). Series 3 of the d_s values was again used. The power curves are shown for a random and balanced arrangement using the 5 % significance level for u; with $k=6$, $n=8$ we have $u_{0\cdot05}=2\cdot485$. The balanced curve crosses the random one rather later than in the preceding illustration.

Fig. 6 (Hudson, III, 6). Series 4 of the d_s values was used. In this case Hudson's two random arrangements A and B are compared with his balanced arrangement; in calculating θ from (20) the estimate σ' was taken from the S_2 of random arrangement A. For $k=6$, $n=4$, $u_{0\cdot05}=2\cdot901$. The balanced curve as usual rises very steeply and crosses the random (B) curve when the chance of detection is about 0·59. Owing to the small number of replications the chance of recognizing small treatment differences is in no case great. In the case of random (B), the calculations were also made using the d-series 2 of Table VI; the change in the power curve was very small, the two curves twisting about one another with four

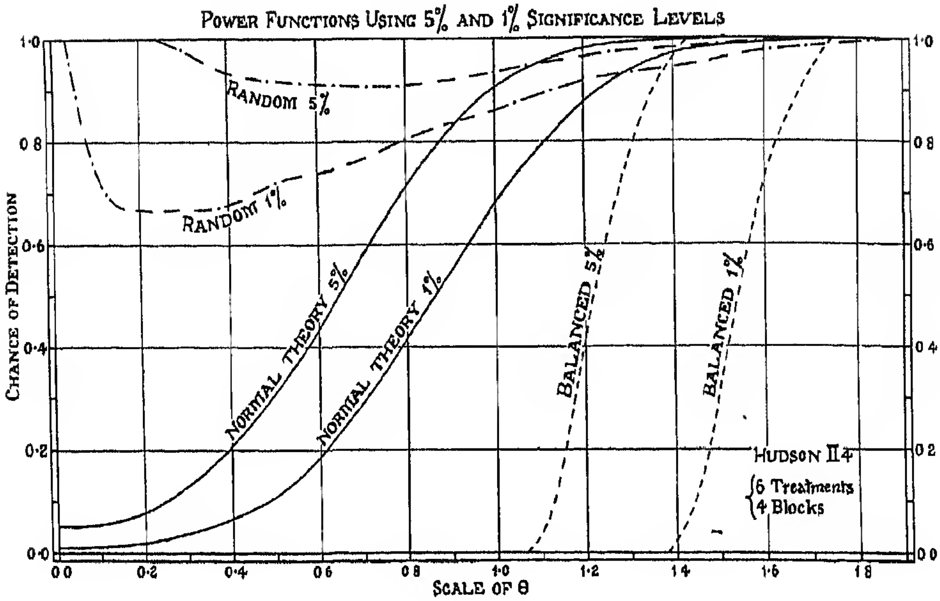


Fig. 3.

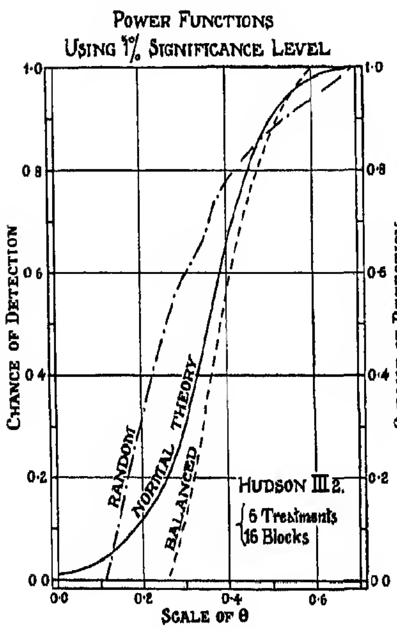


Fig. 4.

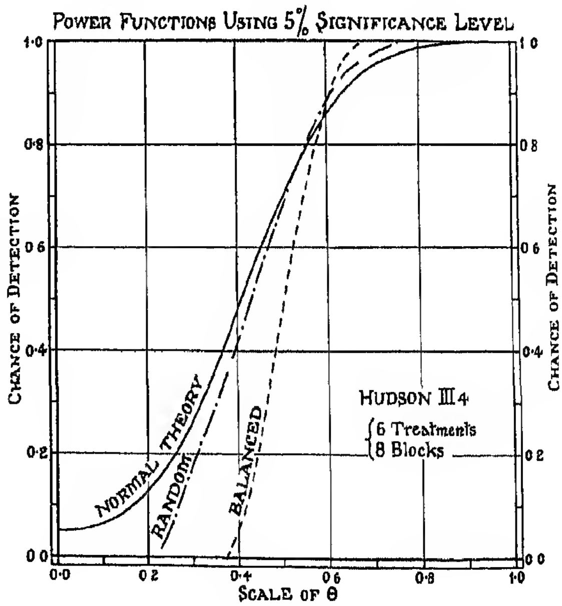


Fig. 5.

points of crossing. This confirms the conclusion suggested in the case of Hudson, II, 4, that it is the standard deviation of the treatment differences, δ_s, not the pattern of their distribution, that really controls the situation.*

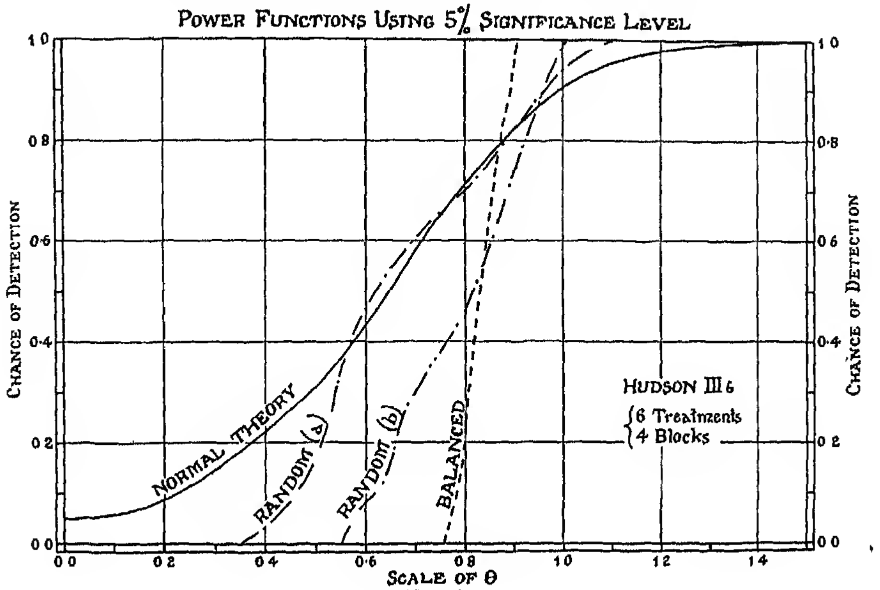


Fig. 6.

5. THE NORMAL THEORY POWER CURVES

These curves have been shown as solid lines in Figs. 1 and 3–6. They are drawn from tabled values given by P. C. Tang (1938). Dr Tang's work is of general application to all analysis of variance problems. In the case of randomized blocks, his results are based on the following assumptions:

(1) The plot yield y_{is} consists of the following additive parts,

$$y_{is} = \beta_i + \delta_s + v_{is}. \qquad \text{......(21)}$$

(2) In this equation β_i is a term constant for the ith block ($i = 1, \ldots, n$), and δ_s is a term constant for the sth treatment ($s = 1, \ldots, k$), subject to the condition $\Sigma(\delta_s) = 0$.

(3) The residuals v_{is} are independent random variables normally distributed about zero with a common standard deviation σ.

Starting from this basis it is possible to obtain the chance that

$$u = \frac{n(n-1)\sum_s (y_{\cdot s} - y_{\cdot\cdot})^2}{\sum_{i,s} (y_{is} - y_{\cdot s} - y_{i\cdot} + y_{\cdot\cdot})^2} \qquad \text{......(22)}$$

will exceed a specified significance level for u, when the δ_s's are in fact not zero.

* This result might be expected having regard to Pitman's (1937a) work regarding the distribution of a correlation coefficient between independent variables under randomization.

Tang's tables give this chance, associated with significance levels $u_{0\,05}$ and $u_{0\,01}$, and suitable values of k and n, for increasing values of

$$\theta = \frac{\sigma_t}{\sigma} = \sqrt{\frac{\Sigma(\delta_t^2)}{k\sigma^2}}. \qquad \ldots\ldots(23)$$

The curves shown in the diagrams have been obtained by plotting this chance of detecting a real difference against θ. Using the 5 % significance level the curve rises from 0·05 at $\theta = 0$ and approaches unity as $\theta \to \infty$. As was pointed out above, when plotting the results obtained from Hudson's data, the true value of σ is unknown and the θ of equation (20), having only an estimate of σ in the denominator, is not strictly comparable with the θ of (23).

Supposing the power curves for a given series of d_s's and for all the $(k!)^{n-1}$ patterns of the randomization set were obtained and superimposed they would form a network of lines.* The normal theory curve would lie somewhere in the centre of these, but how far for a given θ its ordinate would be approximately the average value of the $(k!)^n$ randomization ordinates, I have at present no idea. Since (using $u_{0\cdot05}$) when $\theta = 0$, about 5 % of the randomization set of ordinates will be unity (as for the random arrangement in Fig. 3) and the remaining 95 % will be zero, the average will here agree with the normal theory value of 0·05.

6. Conclusion

The main object of this account has been to make the thesis which "Student" put forward in his last paper as clear as possible with the help of further illustration. In a subject where there are noted differences of opinion, an unambiguous presentation of a case is a first requirement. It seems desirable, however, to end by repeating what appear to be the conclusions which "Student" drew from his discovery regarding the form of the power curves, representing the chance of detection of real treatment differences.

In co-operative experiments undertaken at a number of centres, in which as he emphasized he was chiefly interested, it is of primary concern to study the difference between two (or more) "treatments" under the varying conditions existing in a number of localities. Using a similar notation to that of equations (3) and (4),† in a particular local experiment we shall have for treatments "1" and "2",

$$x_{\cdot 1} = m_{\cdot 1} + \delta_1, \quad x_{\cdot 2} = m_{\cdot 2} + \delta_2, \qquad \ldots\ldots(24)$$

and hence $$x_{\cdot 1} - x_{\cdot 2} = (m_{\cdot 1} - m_{\cdot 2}) + \delta_1 - \delta_2 = m_{\cdot 1} - m_{\cdot 2} + \Delta_{12} \text{ (say)}. \qquad \ldots\ldots(25)$$

The practical problem is then to determine how Δ_{12} varies from one set of conditions to another. For this purpose $m_{\cdot 1} - m_{\cdot 2}$, the difference in "Student's"

* No doubt in their calculation it would be best to take for the estimate of σ, $S_3/n(k-1)$ as previously suggested.

† The notation has been simplified by supposing that $r = s$ and by omitting the brackets round

terminology of treatment error terms within an experiment, should be as small as possible. A balanced arrangement of treatments was in his view more likely to lead to this result than a purely random arrangement. The fact that for such a balanced scheme the calculated estimate of the standard deviation of $m_{.1}-m_{.2}$ would be inaccurate, indeed on the whole somewhat too large, did not worry him; for he considered that the real problem was to compare $x_{.1}-x_{.2}$, the estimate of Δ_{12}, with its *inter*, not *intra*, locality variation. If the error term $m_{.1}-m_{.2}$ was in fact small, although its standard deviation might not be precisely known, values of $x_{.1}-x_{.2}$ would be obtained leading to a consistent interpretation of the situation. On the other hand while randomization might enable the standard deviation of $m_{.1}-m_{.2}$ to be determined without bias, this result would be of little value if the greater fluctuations in this error term made it impossible to interpret the changes in $x_{.1}-x_{.2}$ from one experiment to another.

This was a definite advantage that seemed to be gained from balancing. What, "Student" asked, was lost? Would the single experiment, considered by itself, be no longer of value? As a result of his investigation he felt satisfied that this would not be the case. The balanced experiment admittedly was less likely to detect small treatment differences than the random, and in this sense was inferior. It would not detect differences at all when there was perhaps a fifty-fifty chance that the randomized experiment would do so. Nevertheless it might be argued with reason that useful conclusions for the practical agriculturalist regarding treatment differences cannot be drawn until they can be based on something approaching certainty; in this region when the risk of error is 1 in 10 or less, corresponding to the upper portions of the curves in Figs. 3–6, the balanced lay-out seemed likely to give a slight advantage.

Finally, another practical point was always at the back of "Student's" mind; the ease with which an experiment could be carried out in the field. His general conclusions were not limited to the case of randomized blocks but might be expected to apply in other forms of design.* The randomized treatment pattern is sometimes extremely difficult to apply with ordinary agricultural implements, and he knew from a wide correspondence how often experimenters were troubled or discouraged by the statement that without randomization, conclusions were invalid. The keynote of his paper may perhaps be summarized as follows: in weighing up the consequences of using a given experimental design and applying

* It will be realized that a balanced randomized block may in some cases correspond to a Latin-square lay-out. For example the plan in Table II above contains two 4×4 Latin squares. When this is so the sum of squares, S_2, can of course be reduced by subtracting from it a row (or column) sum of squares. The question will then arise as to whether certain Latin-square patterns would be preferable to others, on "Student's" thesis. The "knight's move" pattern is for example balanced to a greater extent than a randomly selected pattern will usually be, and as Tedin's (1931) work has shown, gives a smaller treatment sum of squares, S_1, on the average. The reduction is however smaller in this case than for randomized blocks, since as "Student" remarked the Latin-square is not only random but balanced "thus conforming to all the principles of allowed witchcraft".

a statistical test to the results of the experiment, it is of more importance to consider (i) the chance of detecting real differences when they exist than (ii) the risk of concluding that a difference exists when it does not. The term "valid" has commonly been associated with a method which ensures a precise knowledge of this latter risk, but may not the most valid procedure be in fact one which, while giving an upper limit to risk (ii), makes as near certain as possible the detection of the larger and therefore most important differences? Whatever answer to this question is favoured, "Student's" last scientific contribution should be invaluable in forcing on the attention of statisticians and experimenters the questions here at issue.

REFERENCES

EDEN, T. & YATES, F. (1933). *J. Agric. Sci.* **23**, 6–16.
IMMER, F. R. (1932). *J. Agric. Res.* **44**, 649–68.
KALAMKAR, R. J. (1932). *J. Agric. Sci.* **22**, 373–85.
NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**, 1–37.
PEARSON, E. S. (1937). *Biometrika*, **29**, 53–64.
PITMAN, E. J. G. (1937*a*). *J. Roy. Statist. Soc.* Suppl. **4**, 225–32.
—— (1937*b*). *Biometrika*, **29**, 322–35.
"STUDENT" (1937). *Biometrika*, **29**, 363–79.
TANG, P. C. (1938). *Statist. Res. Mem.* **2** (in the Press).
TEDIN, O. (1931). *J. Agric. Sci.* **21**, 191–208.
WELCH, B. L. (1937). *Biometrika*, **29**, 21–52.

APPENDIX

The method of obtaining the data for the curves of Figs. 3–6.

In any given case we start with:

(*a*) The values of S_2 and $m_{.(r)} - m_{.(\cdot)}$ given in Table VII and the values of d_s in Table VI.

(*b*) The 720 product sums $\Sigma(m_{.(r)} - m_{.(\cdot)})\, d_s$ given by applying Dr Comrie's method to the two series of six numbers $m_{(r)} - m_{.(\cdot)}$ and d_s $(r, s = 1, 2, \ldots, 6)$.

(*c*) The relation $\delta_s = d_s \sigma_t / \sigma_d$, where σ_t and σ_d are defined by equation (19).

It is then required to determine how many of the corresponding 720 values of the test criterion u defined in equation (5) will fall beyond the significance levels $u_{0\cdot 05}$ and $u_{0\cdot 01}$, for specified values of σ_t^2.

If we write
$$Q = \Sigma(m_{.(r)} - m_{.(\cdot)})\, d_s, \qquad \ldots\ldots(26)$$
then the inequality
$$u = (n-1)\{S_1 + nk\,\sigma_t^2 + 2n\sigma_t\, Q/\sigma_d\}\, S_2^{-1} > u_\alpha,$$
where $\alpha = 0\cdot 05$ or $0\cdot 01$, may be written
$$Q > \frac{\sigma_d}{2n(n-1)\sigma_t}\{u_\alpha S_2 - (n-1)\, S_1 - nk(n-1)\,\sigma_t^2\}. \qquad \ldots\ldots(27)$$
Given the 720 values of Q printed off on sheets by the Hollerith machine, it was

relatively simple to determine how many of these were greater than specified numerical values obtained by inserting into the right-hand side of equation (27) a suitable series of increasing values for σ_t. For the computation and counting involved I am much indebted to Mr D. J. Bishop of the Department of Statistics at University College.

It should be noted that since there is a finite number of values of Q, the power "curve" is not really continuous, but consists of a series of "steps". Counts were however only made for a limited number of values of σ_t, and the points obtained joined graphically by a smooth curve. When some of the d's in a series have the same value, i.e. for series 1, 2 and 4, the power "curves" will tend to show greater sinuosities than otherwise, since the underlying "steps" will be larger.

A GENERALIZATION OF FISHER'S z TEST

By D. N. LAWLEY, B.A.
School of Agriculture, Cambridge

1. Hotelling (1931) has generalized "Student's" t distribution for the case of a normal multivariate population and found the distribution of T, where T is defined by the relation

$$T^2 = \frac{\begin{vmatrix} 0 & \xi_1 & \xi_2 & \cdots & \xi_p \\ \xi_1 & a_{11} & a_{12} & \cdots & a_{1p} \\ \xi_2 & a_{21} & a_{22} & \cdots & a_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \xi_p & a_{p1} & a_{p2} & \cdots & a_{pp} \end{vmatrix}}{\|a_{ij}\|} = \frac{A_{ij}\xi_i\xi_j}{|A|}$$

(i, j being summed over the values $1, 2, \ldots, p$), where $\xi_1, \xi_2, \ldots, \xi_p$ and $x_1^{(s)}, x_2^{(s)}, \ldots, x_p^{(s)}$ ($s = 1, 2, \ldots, n$), are $(n+1)$ independent sets of observations of the variates $x_1, x_2, \ldots, x_p$ which are distributed in a normal multivariate frequency distribution with zero means

$$a_{ij} = \frac{1}{n}\sum_{s=1}^{n} x_i^{(s)} x_j^{(s)},$$

and A_{ij} is the cofactor of a_{ij} in the determinant $|A| = \|a_{ij}\|$.

At a later date Wilks (1932) defined a generalized variance and found the appropriate λ-criteria for testing certain hypotheses concerning the means, variances, and covariances of k normal multivariate populations from which k independent samples have been drawn. These criteria were developed more fully by Pearson and Wilks in a subsequent paper (1933) for the case of two variates, but though the sampling distributions obtained were in certain cases relatively simple the arithmetical calculations required for practical application were, in general, not of a very simple nature.*

In this paper we shall find a quantity which may be regarded as a generalization of Fisher's z and which provides a test suitable for dealing with certain generalized analysis of variance problems, having the advantage of being easy to apply.

* The actual calculation of the λ-criterion which appears to be appropriate in the present problem is no longer than the calculation of v^2 defined below. But when more than two variates are dealt with, only the sampling moments of the λ-criterion seem to be known and the calculation of these certainly involves considerable arithmetic. It is hoped that a fuller comparison of these tests may be made in a further contribution to this *Journal* [Ed.].

2. Using a summation convention for i and j* we define

$$v^2 = \frac{a_{ij} A'_{ij}}{|A'|},$$

where $$a_{ij} = \frac{1}{n_1} \sum_{r=1}^{n_1} x_i^{(r)} x_j^{(r)}, \quad a'_{ij} = \frac{1}{n_2} \sum_{s=1}^{n_2} x_i'^{(s)} x_j'^{(s)},$$

and $$x_i^{(r)} \; r = 1, 2, \ldots, n_1, \quad x_i'^{(s)} \; s = 1, 2, \ldots, n_2$$

represent two independent samples containing respectively n_1 and n_2 sets of values of the variates x_i $(i = 1, 2, \ldots, p)$ which are distributed in a normal multivariate frequency distribution with zero means. We shall find the distribution of v^2.

We may suppose that the variates x_i have zero correlations and equal variances, as otherwise they may always be replaced by linear functions of the x_i having these properties, and v^2 remains invariant under linear transformations.

First consider r to be fixed. Then $(x_i'^{(1)}, x_i'^{(2)}, \ldots, x_i'^{(n_2)}, x_i^{(r)})$ represent the rectangular coordinates of p points P_i in a space V_{n_2+1} of (n_2+1) dimensions, O being the origin and $OX'^{(1)}, \ldots, OX'^{(n_2)}, OX$ being the coordinate axes.

Hotelling uses the result that

$$T_r^2 = \frac{A'_{ij} x_i^{(r)} x_j^{(r)}}{|A'|} = n_2 \cot^2 \phi_r,$$

where ϕ_r is the angle made by OX with the flat space V_p contained by the p lines OP_i.

Let V_{n_2-p+1} be the flat space containing all lines through O perpendicular to V_p, then clearly $\cot \phi_r = \tan \theta_r$, where θ_r is the angle made by OX with V_{n_2-p+1}. Thus $T_r^2 = n_2 \tan^2 \theta_r$.

As the quantities $x_i'^{(1)}, \ldots, x_i'^{(n_2)}, x_i^{(r)}$ vary so do the points P_i, but they are distributed about O with spherical symmetry.

We may consider the space V_{n_2-p+1} to remain fixed while the system of axes varies, then the axis OX moves in such a way that all directions of OX in V_{n_2+1} are equally likely.

Now let V_{n_2} be the flat space contained by the axes $OX'^{(1)}, \ldots, OX'^{(n_2)}$. The intersection of V_{n_2-p+1} with V_{n_2} is a flat space V_{n_2-p} of (n_2-p) dimensions.

It is clear that V_{n_2-p+1} is given by the p equations

$$\sum_{s=1}^{n_2} x_i'^{(s)} X'^{(s)} + x_i^{(r)} X = 0 \quad (i = 1, 2, \ldots, p), \qquad \ldots\ldots(1)$$

and that V_{n_2-p} is given by the $(p+1)$ equations

$$\sum_{s=1}^{n_2} x_i'^{(s)} X'^{(s)} = 0 \quad \text{and} \quad X = 0. \qquad \ldots\ldots(2)$$

Now consider the quantities $x_i'^{(1)}, \ldots, x_i'^{(n_2)}$ as fixed while r takes the values $1, 2, \ldots, n_1$.

* I.e. when the letters i or j appear twice, they are to be summed for all values.

Then the space V_{n_2-p+1} will alter for different values of r but V_{n_2-p} remains unaltered, since equations (2) do not involve the quantities $x_i^{(r)}$. Thus when r takes the values $1, 2, \ldots, n_1$, V_{n_2-p+1} rotates about the fixed space V_{n_2-p}.

We may consider V_{n_2-p+1} to remain fixed and the system of axes to rotate about V_{n_2-p}, then the projection of the axis OX on V_{n_2-p+1} will remain fixed for all successive positions $OX^{(1)}, OX^{(2)}, \ldots, OX^{(n_1)}$ of OX since this is the line through O in V_{n_2-p+1} which is perpendicular to V_{n_2-p}.

As the quantities $x_i^{(r)}$ vary the n_1 lines $OX^{(r)}$ vary independently in the space consisting of all lines through O perpendicular to V_{n_2-p}.

Let θ_r be the angle between $OX^{(r)}$ and V_{n_2-p+1}. Then

$$T_r^2 = \frac{A'_{ij} x_i^{(r)} x_j^{(r)}}{|A'|} = n_2 \tan^2 \theta_r.$$

Hence
$$v^2 = \frac{A'_{ij} a_{ij}}{|A'|} = \frac{1}{n_1} \sum_{r=1}^{n_1} \frac{A'_{ij} x_i^{(r)} x_j^{(r)}}{|A'|} = \frac{n_2}{n_1} \sum_r (\tan^2 \theta_r).$$

Hence we have shown that if $l_1, l_2, \ldots, l_n$ are n_1 lines through O which vary such that all directions in V_{n_2+1} are equally likely, and are independent except for the restriction that they all have the same projection m on V_{n_2-p+1}, then v^2 is distributed as is $\frac{n_2}{n_1} \sum_r (\tan^2 \theta_r)$, where θ_r is the angle between l_r and V_{n_2-p+1}.

The distribution of $\sum_r (\tan^2 \theta_r)$ will be unaltered if instead of the lines l_r lying in the same space V_{n_2+1} we suppose that they are in different spaces $V_{n_2+1}^{(r)}$, each of (n_2+1) dimensions, which intersect in the space V_{n_2-p+1}.

We take rectangular axes $OY_i^{(r)}$ ($i = 1, 2, \ldots, p$ and $r = 1, 2, \ldots, n_1$) and OZ_1 $OZ_2, \ldots, OZ_{n_2-p+1}$ containing a space V of $(n_1 p + n_2 - p + 1)$ dimensions:

$V_{n_2+1}^{(r)}$ is the space contained by the (n_2+1) axes $OY_1^{(r)}, \ldots, OY_p^{(r)}, OZ_1, \ldots, OZ_{n_2-p+1}$,

V_{n_2-p+1} is the space contained by the $(n_2 - p + 1)$ axes $OZ_1, \ldots, OZ_{n_2-p+1}$.

As the lines l_r have the same projection m on V_{n_2-p+1} we may regard them as the projections on the spaces $V_{n_2+1}^{(r)}$ of a line l through O which varies in such a way that all directions of l in V are equally likely.

Now if θ is the angle between l and V_{n_2-p+1} it may be easily shown that

$$\sum_r (\tan^2 \theta_r) = \tan^2 \theta.$$

Hence v^2 is distributed as is $n_2/n_1 \tan^2 \theta$.

It is also easy to prove* that θ is distributed according to the frequency law $f(\theta)\, d\theta = \text{constant} \times \sin^{n_1 p - 1} \theta \cos^{n_2 - p} \theta \, d\theta$. Hence if we put

$$Z = \tfrac{1}{2} \log \left\{ \frac{(n_2 - p + 1)}{n_1 p} \tan^2 \theta \right\} = \tfrac{1}{2} \log \left\{ \frac{(n_2 - p + 1)}{n_2} \times \frac{a_{ij} A'_{ij}}{p\,|A'|} \right\},$$

* For method of proof see Hotelling (1931).

then Z is distributed in Fisher's z distribution with degrees of freedom N_1 and N_2, where $N_1 = n_1 p$ and $N_2 = n_2 - p + 1$.

When $n_1 = 1$ we have the case of Hotelling's T distribution.

When $p = 1$ we obtain the ordinary z distribution of Fisher with degrees of freedom n_1 and n_2.

When $n_2, = \infty$, v^2 is distributed as χ^2/n_1, where χ^2 has $n_1 p$ degrees of freedom, and takes the form $\alpha_{ij} a_{ij}$, where $\alpha_{ij} = C_{ij}/|C|$, $c_{ij} = E(x_i x_j)$, and C_{ij} is the co-factor of c_{ij} in the determinant $|C| = ||c_{ij}||$. The distribution of $\alpha_{ij} a_{ij}$ may easily be obtained directly by a different method.

$\alpha_{ij} a_{ij}$ gives a measure of the "scatter" of the n_1 points whose rectangular coordinates in a space of p dimensions are $(x_1^{(r)}, x_2^{(r)}, \ldots, x_p^{(r)})$ $(r = 1, 2, \ldots, n_1)$, and if the parameters $\{c_{ij}\}$ of the population are known its distribution may be used to test whether the scatter of this set of points, which represent the given sample of size n_1, is too large to be consistent with the hypothesis that the sample has been drawn from a normal multivariate population with zero means and parameters $\{c_{ij}\}$. Usually however the parameters are unknown, in which case the quantities α_{ij} must be replaced by estimates $A'_{ij}/|A'|$ calculated from a second sample, and the distribution of Z may then be used to test whether $v^2 = \dfrac{a_{ij} A'_{ij}}{|A'|}$ is too large to be consistent with the hypothesis that both samples have been drawn from the same normal population.

3. We shall show how the Z test may be applied to experiments in which a number of different treatments are to be compared by analysis of variance methods, and where several variates have been measured.

We suppose that two independent sets of estimates of the variances and covariances have been calculated in the usual manner; one set $\{a_{ij}\}$ having been obtained from the treatment totals and the other set $\{a'_{ij}\}$ from the error differences. Then it may be shown that if there are n_1 degrees of freedom for treatments, and n_2 degrees of freedom for error, and we assume the null hypothesis that there is no effect due to treatments, then

$$a_{ij} = \frac{1}{n_1} \sum_{r=1}^{n_1} x_i^{(r)} x_j^{(r)} \quad \text{and} \quad a'_{ij} = \frac{1}{n_2} \sum_{s=1}^{n_2} x_i'^{(s)} x_j'^{(s)},$$

where $\quad x_i^{(r)}\ (r = 1, 2, \ldots, n_1) \quad$ and $\quad x_i'^{(s)}\ (s = 1, 2, \ldots, n_2)$

represent $(n_1 + n_2)$ independent sets of values of variates x_i $(i = 1, 2, \ldots, p)$ which are distributed in a normal multivariate distribution with zero means.

Hence if we put

$$Z = \tfrac{1}{2} \log \left\{ \frac{(n_2 - p + 1)}{n_2} \times \frac{a_{ij} A'_{ij}}{p\,|A'|} \right\}$$

as before, then Z is distributed according to the distribution obtained in § 2. It will be noticed that the form of Z is not symmetrical in the two sets $\{a_{ij}\}$ and $\{a'_{ij}\}$ so that the two must not be interchanged.

If Z is found to be significantly large it will mean that the set of points whose

coordinates are the sets of treatment means (each point representing a different treatment) are more scattered than would be expected on the null hypothesis that the treatments had no effect, i.e. that there are significant differences between treatments. Of course, in what precedes, for the word "treatments" we may equally well substitute "blocks" such as are used in a randomized block experiment, or "rows" and "columns" of a Latin square.

We give a simple example of the application of the Z test for the case where $p = 2$.

The following figures are taken from an analysis of variance and covariance of Stand (x) and Yield (y) of sugar beet, given by Snedecor (1937, p. 236):

	D.F.	(x^2)	(xy)	(y^2)
Blocks	5	7472·57	−116·56	6·3134
Error	30	28665·10	682·20	23·2326

We shall test whether there are significant differences between blocks.

Carrying out the ordinary analysis of variance test for the variate x we find

$$z = \tfrac{1}{2}\log\left\{\frac{7472{\cdot}57/5}{28665{\cdot}10/30}\right\} = 0{\cdot}2236,$$

with degrees of freedom 5 and 30.

Similarly for y we find

$$z = \tfrac{1}{2}\log\left\{\frac{6{\cdot}3134/5}{23{\cdot}2326/30}\right\} = 0{\cdot}2446$$

also with degrees of freedom 5 and 30.

The 5 % significance point of z for these degrees of freedom is 0·4648, hence neither of the separate analyses of variance of x and y reveal any significant differences between blocks.

But let us put

$$a_{11} = \frac{7472{\cdot}57}{5}, \quad a_{12} = a_{21} = \frac{-116{\cdot}56}{5}, \quad a_{22} = \frac{6{\cdot}3134}{5},$$

and
$$a'_{11} = \frac{28665{\cdot}10}{30}, \quad a'_{12} = a'_{21} = \frac{682{\cdot}20}{30}, \quad a'_{22} = \frac{23{\cdot}2326}{30}.$$

Then
$$\frac{a_{ij}A'_{ij}}{p\,|A'|} = 7{\cdot}683 \qquad (p=2;\ i,\ j=1,2),$$

$n_1 = 5$ and $n_2 = 30$.

Hence
$$Z = \tfrac{1}{2}\log\left(\tfrac{29}{30}\times 7{\cdot}683\right)$$
$$= 1{\cdot}0026,$$

$N_1 = n_1 p = 10$ and $N_2 = n_2 - p + 1 = 29$.

The value of Z obtained is significantly large even at the 0·1 % significance level, as for degrees of freedom 10 and 29 the 0·1 % point of z is 0·7283, thus the differences between blocks are shown to be strongly significant.

The largeness of Z is partly accounted for by the fact that the "between blocks" correlation coefficient (r) of $-0{\cdot}537$ differs greatly from the error correlation coefficient (r') of $+0{\cdot}836$, and also partly by the fact that due to r' being large $|A'|$ is small.

4. Wilks (1934) has considered the regression of a set of dependent variates upon another set of independent variates. We shall now show how the distribution of Z may be used to test the significance of the composite regression of the dependent variates $\{y_i\}$ $(i=1, 2, \ldots, p)$ on the independent variates $\{x_r\}$ $(r=1, 2, \ldots, m)$.

In what follows we use a summation convention for all lower suffixes.

Let us suppose that the y's and x's are deviations from means and that the variate y_i is normally distributed about the linear regression function $\beta_{ir}x_r$ so that the joint probability law of $\{y_i\}$ is

$$\pi^{-p/2}\,||\,\alpha_{ij}\,||^{\frac{1}{2}}\,e^{-\frac{1}{2}\alpha_{ij}(y_i-\beta_{ir}x_r)(y_j-\beta_{js}x_s)}\,dy,$$

where $\alpha_{ij} = \dfrac{C_{ij}}{|C|}$ and $c_{ij} = E(y_i y_j)$ for fixed $\{x_r\}$ and $dy = \prod\limits_{i=1}^{p}(dy_i)$.

The coefficient β_{ir} is defined to be the generalized regression coefficient of y_i on x_r.

Now let $\{y_i^{(\alpha)}\}$ and $\{x_r^{(\alpha)}\}$ where $\alpha = 1, 2, \ldots, n$ represent a sample of size n from the given population, where we suppose that the $y_i^{(\alpha)}$ and $x_r^{(\alpha)}$ are deviations from the sample means, so that

$$\sum_{\alpha=1}^{n} y_i^{(\alpha)} = 0 \quad (i=1, 2, \ldots, p) \qquad \text{and} \qquad \sum_{\alpha=1}^{n} x_r^{(\alpha)} = 0 \quad (r=1, 2, \ldots, m).$$

We fit lines of the form $Y_i = b_{ir}x_r$ to the given data by choosing the coefficients b_{ir}, which are estimates of the β_{ir}, so as to make the quantity

$$\sum_{\alpha}\{\alpha_{ij}(y_i^{(\alpha)} - Y_i^{(\alpha)})(y_j^{(\alpha)} - Y_j^{(\alpha)})\},$$

where $Y_i^{(\alpha)} = b_{ir}x_r^{(\alpha)}$, a minimum.

This expression is a minimum when $\{b_{ir}\}$ satisfy the following mp equations

$$\sum_{\alpha}\{\alpha_{ij}x_r^{(\alpha)}(y_j^{(\alpha)} - b_{js}x_s^{(\alpha)})\} = 0,$$

i.e.

$$\alpha_{ij}(d_{rj} - g_{rs}b_{js}) = 0,$$

where

$$g_{rs} = \sum_{\alpha} x_r^{(\alpha)}x_s^{(\alpha)} \quad \text{and} \quad d_{jr} = \sum_{\alpha} y_j^{(\alpha)}x_r^{(\alpha)}.$$

These equations are satisfied when $g_{rs}b_{js} = d_{jr}$. Hence

$$b_{ir} = \frac{G_{rs}}{|G|}d_{is},$$

where G_{rs} is the cofactor of g_{rs} in the determinant $|G| = ||\,g_{rs}\,||$. The quantities

b_{ir}, being weighted sums of the $y_i^{(\alpha)}$, are distributed normally and it is easy to show that

(1) $E(b_{ir}) = \beta_{ir}$;

(2) The variance of b_{ir} is $\dfrac{G_{rr}}{|G|} \cdot c_{ii}$ (i and r not summed);

(3) The covariance of b_{ir} and b_{js} is $\dfrac{G_{rs}}{|G|} \cdot c_{ij}$.

Now
$$\begin{aligned}
&\sum_\alpha (y_i^{(\alpha)} - Y_i^{(\alpha)})(y_j^{(\alpha)} - Y_j^{(\alpha)}) \\
&= \sum_\alpha (y_i^{(\alpha)} - b_{ir}x_r^{(\alpha)})(y_j^{(\alpha)} - b_{js}x_s^{(\alpha)}) \\
&= \sum_\alpha y_i^{(\alpha)} y_j^{(\alpha)} - b_{ir}\Sigma y_j^{(\alpha)} x_r^{(\alpha)} - b_{js}\Sigma y_i^{(\alpha)} x_s^{(\alpha)} + b_{ir}b_{js}\Sigma x_r^{(\alpha)} x_s^{(\alpha)} \\
&= \sum_\alpha y_i^{(\alpha)} y_j^{(\alpha)} - \frac{G_{rs}}{|G|} d_{ir} d_{js}.
\end{aligned}$$

Hence
$$\sum_\alpha y_i^{(\alpha)} y_j^{(\alpha)} = \sum_\alpha (y_i^{(\alpha)} - Y_i^{(\alpha)})(y_j^{(\alpha)} - Y_j^{(\alpha)}) + b_{is}\sum_\alpha y_j^{(\alpha)} x_s^{(\alpha)}.$$

Let us assume the null hypothesis that the regression coefficients β_{ir} are all zero. Then it may be proved that for each i we can find $(n-1)$ linear functions $\xi_i^{(\gamma)}$ $(\gamma = 1, 2, \ldots, n-1)$ of the $y_i^{(\alpha)}$ which are independently and normally distributed with equal variances and zero means, and which are such that

$$b_{is}\sum_{\alpha=1}^{n} y_j^{(\alpha)} x_s^{(\alpha)} = \sum_{\gamma=1}^{m} \xi_i^{(\gamma)} \xi_j^{(\gamma)},$$

$$\sum_{\alpha=1}^{n} (y_i^{(\alpha)} - Y_i^{(\alpha)})(y_j^{(\alpha)} - Y_j^{(\alpha)}) = \sum_{\gamma=m+1}^{n-1} \xi_i^{(\gamma)} \xi_j^{(\gamma)}$$

and
$$\sum_{\alpha=1}^{n} y_i^{(\alpha)} y_j^{(\alpha)} = \sum_{\gamma=1}^{n-1} \xi_i^{(\gamma)} \xi_j^{(\gamma)}.$$

Thus
$$a_{ij} = \frac{1}{m} b_{is} \sum_{\alpha=1}^{n} y_j^{(\alpha)} x_s^{(\alpha)}$$

and
$$a'_{ij} = \frac{1}{(n-m-1)} \sum_{\alpha=1}^{n} (y_i^{(\alpha)} - Y_i^{(\alpha)})(y_j^{(\alpha)} - Y_j^{(\alpha)})$$

are independently distributed estimates of c_{ij} with degrees of freedom m and $(n-m-1)$ respectively. Hence if we put

$$Z = \tfrac{1}{2}\log\left\{\frac{(n-m-p)}{(n-m-1)} \times \frac{a_{ij}A'_{ij}}{p|A'|}\right\},$$

then Z is distributed in Fisher's z distribution with degrees of freedom N_1 and N_2, where $N_1 = mp$ and $N_2 = (n-m-p)$.

This gives the required test of significance of the regression of $\{y_i\}$ on $\{x_r\}$. If the value of Z obtained reaches the level of significance then the null hypothesis is considered to be disproved.

REFERENCES

HOTELLING, H. (1931). "The generalization of 'Student's' ratio." *Ann. Math. Statist.* **2**, 359–78.

PEARSON, E. S. & WILKS, S. S. (1933). "Methods of statistical analysis appropriate for k samples of two variables." *Biometrika*, **25**, 353–78.

SNEDECOR, G. W. (1937). *Statistical Methods*. Iowa: Collegiate Press, Inc., Ames.

WILKS, S. S. (1932). "Certain generalizations in the analysis of variance." *Biometrika*, **24**, 471–94.

—— (1934). "Moment generating operators for determinants of product moments in samples from a normal system." *Ann. Math.* **35**, 312–40.

MISCELLANEA

(i) Applicability of the z Test to a Poisson Distribution

By R. A. CHAPMAN

Assistant Silviculturist, Southern Forest Experiment Station

THE distribution of z was derived on the assumption that the parent population was normally distributed; in actual practice this assumption usually does not hold. Several investigators have studied the distribution of z or its related function η^2* obtained from non-normal populations. Pearson (1931) studied the distribution of η^2 for several types of population, in which the values for β_1 and β_2 were (0·0, 2·5), (0·0, 4·1), (0·0, 7·1), (0·2, 3·3), (0·5, 3·7), and (1·0, 3·8). The results of these experiments suggest that within the range of the experimental populations tried, the assumption of normality gives satisfactory results for most work. Eden & Yates (1933) also did some sampling work, using height measurements of wheat, but they did not deal with a very skewed distribution. At the Southern Forest Experiment Station the author recently had occasion to draw experimental samples from a distribution even more skewed than that used by Pearson or by Eden and Yates. This paper presents a brief report of the results obtained.

TABLE I

Original population sampled, and range of numbers used in sampling

Value of variable	Number of occurrences	Range of numbers
0	368	000–367
1	368	368–735
2	184	736–919
3	61	920–980
4	15	981–995
5	3	996–998
6	1	999

The parent population sampled was a Poisson distribution with a mean equal to 1, $\beta_1 = 1{\cdot}0401$, $\beta_2 = 4{\cdot}1031$.† This form of distribution was found in a study of the effect of greenhouse treatments on the mortality of longleaf and slash pine seedlings. The actual distribution sampled is shown in column 2 of Table I. From this population 100 samples of 16 values each were drawn, with the aid of Tippett's random numbers, using the last three of the four digits in a column. Each 3-digit number drawn was then referred to the class interval shown in column 3 of Table I. If the number was between 000 and 367 it was called 0; if it was between 368 and 735 it was called 1; and so on. As the numbers (samples) were drawn they were separated into sub-samples of four values each. For each

* $\eta^2 = \underset{x}{S}[n_x(\bar{Y}_x - \bar{Y})^2]/S[(Y-\bar{Y})^2]$.

† Due to some small discrepancy in the frequency distribution, the values of β_1 and β differed from their theoretical values of 1 and 4, respectively.

TABLE II

Actual and theoretical frequency of F with χ^2 test

Class interval of F (central values)	Actual frequency $= a$	Theoretical frequency $= t$	$(a-t)$	$\frac{(a-t)^2}{t}$
0·2	22	24·4	−2·4	0·2361
0·6	33	23·8	+9·2	3·5563
1·0	17	16·6	+0·4	0·0096
1·4	8	11·0	−3·0	0·8182
1·8	4	7·3	−3·3	1·4918
2·2	4	4·8	−0·8	0·1333
2·6	4	3·6	+0·4	0·0444
3·0	2	2·1		
3·4	1	1·8		
3·8	2 } 8	1·1 } 8·5	−0·5	0·0294
Above 4·0	3	3·5		
Total	100	100·0		$6{\cdot}3191 = \chi^2$

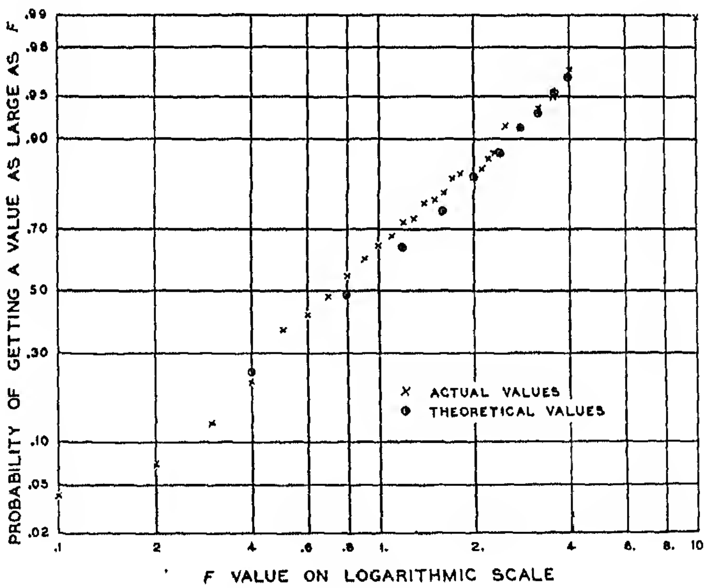


Fig. 1.

of the samples the total sum of the squares was divided into two parts: between means of sub-samples, with three degrees of freedom; and within sub-samples, with 12 degrees of freedom. Using the two estimates of variance so derived, an F value was computed:

$$F = \frac{\sigma_B^2}{\sigma_W^2}.$$

This is the ratio of the variance between means of sub-samples to the variance within sub-samples. For convenience of computation the results are presented as F values rather than as z values.

The actual distribution of F from these 100 samples is shown in Table II, column 2, and in Fig. 1. The theoretical distribution of F is also shown.

The Chi-square test of the comparison of actual and theoretical frequencies in Table II gives a χ^2 of 6·3191 which, with 7 degrees of freedom, has a probability of about 0·5. The agreement between the actual and theoretical distributions is therefore satisfactory, and this result confirms the conclusion reached by others that the z test is applicable to skewed distributions.

REFERENCES

PEARSON, EGON S. (1931). "The analysis of variance in cases of non-normal variation." *Biometrika*, **23**, 114–33.

EDEN, T. & YATES, F. (1933). "On the validity of Fisher's z test when applied to an actual example of non-normal data." *J. Agric. Sci.* **33**, 6–17.

(ii) The Distribution of the Ratio of Estimates of the Two Variances in a Sample from a Normal Bi-variate Population

BY D. J. FINNEY, B.A., *Clare College, Cambridge*

THE distribution of this ratio has been investigated by Bose (1935) by a method dependent on term-by-term integration of infinite series. The simplicity of his result suggested the possibility of the more direct approach given below, which is followed by the evaluation of the probability integral for the distribution. It is then shown how, by a simple transformation of existing tables, a test of significance may be applied when the population correlation coefficient is known, and how the test may be adapted when only a sample estimate of this correlation is available.

By a suitable choice of scales, any normal bi-variate distribution function may be written as

$$F(x_1, x_2) = \frac{1}{2\pi(1-\rho^2)^{\frac{1}{2}}} e^{-\frac{1}{2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}$$

The three second-order sums for a sample of size n from this population may be defined as

$$c_{ij} = \sum_{p=1}^{n} (x_{ip} - \bar{x}_i)(x_{jp} - \bar{x}_j), \quad i, j = 1, 2,$$

where (x_{1p}, x_{2p}) are corresponding pairs of observations. The distribution of the c_{ij} is $V(c_{11}, c_{22}, c_{12})\, dc_{11} dc_{22} dc_{12}$, where

$$V(c_{11}, c_{22}, c_{12}) = K(c_{11}c_{22} - c_{12}^2)^{\frac{n-4}{2}} e^{-\frac{1}{2(1-\rho^2)}(c_{11} - 2\rho c_{12} + c_{22})}$$

and

$$K^{-1} = \pi^{\frac{1}{2}} 2^{n-1} (1-\rho^2)^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{n-2}{2}\right).$$

If s_1^2, s_2^2 are the estimates of the variances of x_1 and x_2 and r is the estimate of ρ from this sample, then

$$s_1^2 = \frac{c_{11}}{n-1}, \qquad s_2^2 = \frac{c_{22}}{n-1}, \qquad r = \frac{c_{12}}{(c_{11}c_{22})^{\frac{1}{2}}}.$$

It follows that the distribution of $\omega = s_1/s_2$ is $V(\omega)\, d\omega$, where

$$\begin{aligned} V(\omega) &= 2K\omega^{n-2} \int_{-1}^{1} dr \int_{0}^{\infty} dc_{22} (1-r^2)^{\frac{n-4}{2}} c_{22}^{n-2} e^{-\frac{c_{22}}{2(1-\rho^2)}(1 - 2\rho\omega r + \omega^2)} \\ &= 2K'\omega^{n-2} \int_{-1}^{1} dr (1-r^2)^{\frac{n-4}{2}} (1 - 2\rho\omega r + \omega^2)^{-(n-1)} \end{aligned}$$

with

$$K' = K\Gamma(n-1) 2^{n-1} (1-\rho^2)^{n-1}.$$

The substitution
$$r = \frac{(\lambda+\mu)v-(\lambda-\mu)(1-v)}{(\lambda+\mu)v+(\lambda-\mu)(1-v)},$$
where $\lambda = 1+\omega^2$, $\mu = 2\rho\omega$, reduces the integral to the form

$$V(\omega) = 2^{n-2}K'\omega^{n-2}(\lambda^2-\mu^2)^{-\frac{n}{2}}\int_0^1\left\{(\lambda+\mu)v^{\frac{n-2}{2}}(1-v)^{\frac{n-4}{2}}+(\lambda-\mu)v^{\frac{n-4}{2}}(1-v)^{\frac{n-2}{2}}\right\}dv$$
$$= \frac{2(1-\rho^2)^{\frac{n-1}{2}}}{B\left(\frac{n-1}{2},\frac{n-1}{2}\right)}\cdot\frac{\omega^{n-2}}{(1+\omega^2)^{n-1}}\left\{1-\frac{4\rho^2\omega^2}{(1+\omega^2)^2}\right\}^{-\frac{n}{2}},$$

which is the result given by Bose. If the population values of the variances are σ_1^2, σ_2^2 the same distribution holds for
$$\omega = \frac{s_1}{\sigma_1}\Big/\frac{s_2}{\sigma_2}.$$

When $\rho = 0$, the distribution reduces to
$$V(\omega) = \frac{2}{B\left(\frac{n-1}{2},\frac{n-1}{2}\right)}\cdot\frac{\omega^{n-2}}{(1+\omega^2)^{n-1}},$$

which is a particular case of that obtained by Fisher for the ratio of two independent estimates of a variance.

Now the distributions of ω and ω^{-1} are identical—as is otherwise obvious from the definition of ω. Thus a sample value of the ratio may be so chosen as to give $\omega \geqslant 1$, and the probability of obtaining $\omega \geqslant \Omega \geqslant 1$ by random sampling is then

$$P(\Omega) = \int_\Omega^\infty V(\omega)\,d\omega$$
$$\propto \int_\Omega^\infty \left(\frac{\omega}{1+\omega^2}\right)^{n-1}\left\{1-4\rho^2\left(\frac{\omega}{1+\omega^2}\right)^2\right\}^{-\frac{n}{2}}\frac{d\omega}{\omega}.$$

The substitution $\qquad \omega^2+\omega^{-2} = e^{2z}+e^{-2z}-\rho^2(e^{2z}+e^{-2z}-2)$

transforms this to $\qquad P(\Omega) \propto \int_Z^\infty \frac{e^{(n-1)z}}{(1+e^{2z})^{n-1}}dz,$

which is the probability integral of Fisher's z-distribution with degrees of freedom $n_1 = n_2 = n-1$, whence it follows that
$$P(\Omega) = I_x\left(\frac{n-1}{2},\frac{n-1}{2}\right),$$
where
$$x = \frac{1}{2}\left\{1-\frac{\Omega-\Omega^{-1}}{\sqrt{\{(\Omega+\Omega^{-1})^2-4\rho^2\}}}\right\},$$

and the probability integral can be read from tables of the Incomplete Beta Function. Alternatively, significance levels can be constructed by entering Fisher's z-table with $n_1 = n_2 = n-1$ and, when ρ is known, finding the value of Ω corresponding to the Z so obtained. Thus with $n = 5$, for various values of ρ the 5 % and 1 % points of Ω^2 are as follows:

ρ	0·0	0·1	0·3	0·5	0·7	0·9	0·95	0·99
5 % point	6·388	6·342	5·968	5·217	4·072	2·457	1·923	1·348
1 % point	15·978	15·837	14·710	12·450	9·050	4·443	3·040	1·686

This test for significance can only be applied when the population parameter, ρ, is known. When only a sample estimate of ρ, r say, is available, the method suggested by Hirschfeld (1937) can be adopted. For fixed n and Ω, $P(\Omega)$ is a monotonic function of ρ^2 and $P(\Omega) \to 0$ as $\rho^2 \to 1$. Thus, if Z is the entry of Fisher's table corresponding to $n_1 = n_2 = n-1$ at the chosen level of significance, Ω, determined from the sample, will be significant if $\rho^2 \geqslant P^2$, where

$$P^2 = \frac{e^{2Z} + e^{-2Z} - \Omega^2 - \Omega^{-2}}{e^{2Z} + e^{-2Z} - 2}.$$

Clearly the meaning of a negative value of P^2 would be that significance is obtained by the ordinary z-test ($\Omega^2 > e^{2Z}$) and that therefore Ω is significant whatever the value of ρ may be.

If, when tested by Fisher's transformation, $|r|$ is significantly greater than the critical value $|P|$, the significance of Ω is assured. It is clearly not necessary that r should be determined from the same sample as Ω and it will be advantageous to use a more precise estimate when such is available. When r is small and based on few degrees of freedom there is little hope of attaining significance by this method if the ordinary z-test has failed to show its existence, but for a large r with many degrees of freedom the value of Ω for which significance is reached will be very considerably reduced.

Example. In a paper on "Physical measurements and vital capacity" Mumford & Young (1923) give results of measurements of standing height and stem length for different age groups of schoolboys. Taking all measurements as percentages of their respective means, from Tables I and II of this paper it is found that, for 173 boys aged 13–14 years,

$$\text{Standard deviation of standing height} = 5{\cdot}299\ \%,$$
$$\text{Standard deviation of stem length} = 4{\cdot}766\ \%.$$

Hence $\Omega^2 = 1{\cdot}236.$

Also $r = 0{\cdot}878.$

Using Fisher (1936), § 41, to obtain the 1 % point of Z with $n_1 = n_2 = 172$,

$$e^{2Z} = 1{\cdot}427,$$

and it follows that $P = 0{\cdot}805.$

Transforming the correlations $z_r = 1{\cdot}367,$

$$z_P = 1{\cdot}114,$$

and it is seen that $0{\cdot}253\sqrt{170} = 3{\cdot}30$ is a unit normal deviate significant at the 1 % level. It is thus demonstrated that the stem length is proportionately less variable than the standing height in the population considered.

REFERENCES

Bose, S. (1935). *Sankhyā: Indian J. Statistics*, **1**, 65.
Fisher, R. A. (1936). *Statistical Methods for Research Workers*, 6th ed.
Hirschfeld, H. O. (1937). *Biometrika*, **29**, 65.
Mumford, A. A. & Young, M. (1923). *Biometrika*, **15**, 108.

(iii) Gauss' Quadratic Formula with Twelve Ordinates

By B. de F. BAYLY
Assistant Professor in Electrical Engineering, University of Toronto

J. O. Irwin (1923) has pointed out the desirability of knowing the constants for Gauss' quadrature formula using twelve ordinates. This computation, which is quite laborious, has been completed, and the results are given herewith.

Legendre's polynomial of the twelfth degree is

$$P_{12}(x) = \frac{1}{1024}(676039\,x^{12} - 1939938\,x^{10} \quad 2078505\,x^{8} - 1021020\,x^{6} \quad 225225\,x^{4} - 18018\,x^{2} \quad 231).$$

If this is equated to zero the following values of the roots are obtained:

a_1	−0·9815 6063 4246 7	$\log(-a_1)$	$\bar{1}$·9919 1713 2571 1812 28
a_2	−0·9041 1725 6370 5		$\bar{1}$ 9562 2475 8453 6039 54
a_3	−0·7699 0267 4194 3		$\bar{1}$·8864 3582 8118 1096 56
a_4	−0·5873 1795 4286 6		$\bar{1}$·7688 7327 7411 0133 37
a_5	−0·3678 3149 8998 2		$\bar{1}$·5656 4891 7004 6865 00
a_6	−0·1252 3340 8511 5		$\bar{1}$·0977 2020 1052 5827 95

The remaining roots a_7 to a_{12} are equal to a_6 to a_1, only with positive sign.

The equation for finding an integral is as follows:

$$\int_p^q f(x)\,dx = \sum_{n=1}^{n=12} \frac{q-p}{2} f\left(\frac{q+p}{2} + \frac{q-p}{2}\,a_n\right) b_n,$$

the values of b_n being given in the following table:

b_1 and b_{12}	0·0471 7533 6386 4
b_2 and b_{11}	0·1069 3932 5995 3
b_3 and b_{10}	0·1600 7832 8543 4
b_4 and b_9	0·2031 6742 6723 2
b_5 and b_8	0·2334 9253 6538 4
b_6 and b_7	0·2491 4704 5813 4

As a check on these values $\log_e 2$ was calculated and found correct to the thirteenth place.

In the article referred to above it was suggested that Gauss' method with twelve ordinates would be satisfactory for computing such functions as the incomplete Beta-function. The function*

$$I_{0\cdot6}(16\cdot1,\ 5\cdot2) = \frac{\int_0^{0\cdot6} x^{15\cdot1}(1-x)^{4\cdot2}\,dx}{\int_0^1 x^{15\cdot1}(1-x)^{4\cdot2}\,dx}$$

was computed by this method and the value found was

0·0567 0985 9126,

the correct value being 0·0567 0986 1893.

* Only the numerator was calculated by this method. The denominator of course is obtained from tables of the Gamma-function.

The error apparently is about one part in twenty million. Owing to lack of time however this final check computation has not been very carefully checked. In any event the use of Gauss' method is not recommended for functions of this type as the above computation took several hours even with every possible aid to calculation. It is felt that with functions of this type other quadrature formulae even using three times as many ordinates would be less laborious.

REFERENCE

IRWIN, J. O. (1923). *Tracts for Computers*, edited by Karl Pearson; No. X: "On quadrature and cubature, or on methods of determining approximately single and double integrals." Camb. Univ. Press.

(iv) **Introduction to Mathematical Probability.** By J. V. USPENSKY. London: McGraw-Hill Publishing Company, 1937. Price 30*s*.

THERE are two principles which should be followed by the writer of any elementary text-book. First, the theory should be set out simply and directly, so that it is intelligible to a reader who has no previous knowledge of the subject, and secondly, the theory should be illustrated by a number of worked examples, so that the reader having been shown "why" can understand "how". Prof. Uspensky follows these two principles, and his book should become a model for writers on the theory of probability for many years.

The author gives a clear delineation of the development of the classical theory of probability of to-day. He does not attempt to give its applications to other sciences, but his illustrations are such that the reader will have very little difficulty in finding these applications for himself. An example of this is found in his derivation of the distribution of "Student's" "*t*" using characteristic functions.

At a time when many books on probability are being written, and when the theory of probability is being applied in many different fields, it is very satisfactory to find the theory developed with such absolute clarity and unusual attention to rigour. Much is presented which hitherto has been unattainable except by a study of the literature of the Russian school, but it is possible to learn much also from his treatment of the theorems which are well known; for example, the proof of the famous and much-discussed theorem of Laplace is considerably enhanced by the method of estimating the error involved in its application.

Chapters I and II contain approximately the theory of probability as usually given in text-books on algebra. Chapter III discusses the problem of repeated trials and contains a very ingenious method, due to Markoff, of approximating to large factorials, and the sum of large factorials, by means of continued fractions. Chapter IV is exceptionally valuable to students of the theory of estimation, for the author discusses thoroughly the theorem of Bayes and its applications, and leaves no room for doubt of the fact that its application to practical problems is usually invalid, because of the lack of the necessary data. Perhaps here an example might have been added on its validity when applied to certain problems arising in the Mendelian theory.

Chapter V introduces us to the simple theory of "Markoff chains", and the use of difference equations in solving questions in probability. Cantelli's theorem on the upper limit of a probability is given in Chapter VI, while Chapter VII contains the proof of the theorem of Laplace to which we have already referred. The succeeding chapter on "Further Considerations on Games of Chance" is not important from a point of view of the theory of probability, although it may be read with profit for its ingenious algebra.

In Chapter IX we find the first discussion of a stochastic variable, and the elements of the mathematical theory of expectations are developed so as to lead us easily and naturally

in the next chapter to Tschebysheff's Lemma, the law of large numbers, and Markoff's theorem on the large numbers. The author discusses shortly the "strong law of large numbers", this last being proved as an example at the end of the chapter. These laws are illustrated by numerical examples in both Chapters x and xi.

Chapter xii is headed "Probabilities in Continuum", and is concerned with the definitions of the characteristic function and the distribution function.

Prof. Uspensky states in his preface that these twelve chapters may be read by persons "without advanced mathematical knowledge", while the remaining chapters, incorporating the results of modern researches, require from their readers a "more mature mathematical preparation". While the present writer thinks that the words "without advanced mathematical knowledge" might be qualified, since some of the analysis is by no means easy, there is no doubt that Chapters xiii onwards are unrivalled in any comparable English text-book for the beauty and elegance of their methods of analysis.

Chapter xiii discusses the Stieltjes Integral and its application in the theory of characteristic functions, and Liapounoff's inequality for moments. The examples given require a knowledge of contour integration. The next chapter follows in logical sequence with applications of this theory to further problems. Here we find Liapounoff's theorem stated and proved with the aid of the characteristic function and the Liapounoff inequality.

The remaining two chapters are of interest primarily to statisticians. The bivariate normal surface is discussed with the aid of the previous analysis, and the distributions of several different functions of normally distributed variables are obtained, notably those of s, r and t.

The whole volume is illustrated by a wealth of examples, each of which adds to our understanding of the theory, if not to the theory itself. It is a pleasant surprise and stimulus to find theorems set as an exercise, with the outline of their proof given as an aid. This book is so good that it should remain a classic in the literature of the theory of probability for many years.

One minor point of criticism might be raised. The present writer, at least, finds that the notation used by Prof. Uspensky in the first few chapters is confusing. Consider for example, the theorem on compound probability on p. 31 Prof. Uspensky writes

$$(AB) = (A).(B, A)$$

which is interpreted by the statement "the probability of the simultaneous occurrence of A and B is given by the product of the unconditional probability of the event A, by the conditional probability of B supposing A actually occurred". It seems to the writer that the following notation is less confusing:

$$P\{AB\} = P\{A\}.P\{B|A\},$$

which expression is interpreted in the same way as the above, where $P\{\ \}$ stands for "the probability that". However, notation is merely a matter of taste, and this small point does not detract from the value of the book as a whole.

Prof. Uspensky modestly describes the subject of his book as the *Elementary Theory of Probability*. This raises the hope that one day we shall have another book from his pen in which he will write of the theory of probability based on the theory of measure and Lebesque-Stieltjes integration. Such a book would be read eagerly by all those who have enjoyed this present volume.

F. N. David.

Department of Statistics,
University College, London.

(v) Heterostylism in natural populations of the Primrose, *Primula acaulis*

By J. B. S. HALDANE

The primrose is one of the heterostylic species of *Primula*, the flowers being either "thrum" with short style and anthers at the mouth of the tube, or "pin" with long style and anthers in the tube. It is known that the two forms exist in nature in about equal numbers; and that "legitimate" unions between the two types are much more fertile than "illegitimate" unions within a type (Darwin, 1877). Gregory (1915) found that thrum is dominant to pin, all natural thrums examined being heterozygous, so that the union of thrum and pin gave the two types in almost equal numbers (229 thrum, 236 pin), while thrum selfed gave 3 thrum : 1 pin (39 thrum, 13 pin).

In counting natural populations I had two objects in view, to see whether the ratio of the two types diverged significantly from unity, and whether individual populations varied significantly from the mean proportion. I usually counted between 100 and 200 plants growing as closely together as possible, so that they might be regarded as a naturally inter-crossing population. Most of the populations were found on roadsides in Wales and southern England. Those at Garreg 1 and Port Meirion were in open woods, that at Garreg 2 in a pasture, while those at Ymstyllynn and Pangbourne were by the sides of railways.

A certain difficulty was occasionally experienced in deciding whether two plants growing close together could have arisen by vegetative reproduction from one seedling. Where there was a doubt only one flower was observed, even though further inspection sometimes showed both thrum and pin plants in the same clump.

The results are given in Table I together with those of Darwin (1877), Scott (quoted by Darwin) and v. Tschermak (1923). It may be remarked that v. Tschermak gives Darwin's figures incorrectly. v. Tschermak's sample was from a single locality. It will be seen that the totals in no case diverge significantly from equality. The grand total gives $50 \cdot 83 \pm 0 \cdot 79$ % thrums. Thus if this is the true ratio another 6000 or so plants will have to be counted to establish a probably significant deviation from equality such as de Winton & Haldane (1933) found in experimental crosses of pin × thrum (but not thrum × pin) in *Primula sinensis*. The former cross gave $51 \cdot 45 \pm 0 \cdot 67$ %, the latter $49 \cdot 34 \pm 0 \cdot 60$ % of thrums.

We have next to ask whether it is legitimate to calculate the standard error of this ratio. Can the individual populations be regarded as samples from a single large population? Or are they heterogeneous, even though their sum gives a ratio consistent with equality? The values of χ^2 for an expectation of equality are given in Table I. The total for my data is $\chi^2 = 24 \cdot 93$ with $n = 17$. Using Haldane's (1938) equation (10) we find $P = 0 \cdot 096$. For all the data $\chi^2 = 27 \cdot 02$ for $n = 20$, so $P = 0 \cdot 131$.

If we wish merely to test for homogeneity, with one less degree of freedom, we can use the following transformation (Haldane, 1936*a*).

If c be the true frequency of one class in a $(2 \times n)$-fold table, and c' the assumed value (here $\frac{1}{2}$), if N be the total number of the population, and if χ'^2 be the value of χ^2 found when the value c' is assumed, then the true value is

$$\chi^2 = \frac{1}{c(1-c)} [c'(1-c')\chi'^2 - (c-c')^2 N].$$

For my seventeen populations $c = 0 \cdot 49088$, so $\chi^2 = 24 \cdot 49$, for $n = 16$. Hence $P = 0 \cdot 080$. For all twenty populations $c = 0 \cdot 49146$, $\chi^2 = 25 \cdot 82$, $n = 19$, $P = 0 \cdot 135$. There is thus an indication, but certainly no proof, of heterogeneity. Nevertheless, I am inclined to suspect that larger counts would reveal it. For I got the definite impression that runs of five or

TABLE I

Place	Thrum	Pin	χ^2
Red Roses (Pembroke)	130	92	6·811
Bredenbury 1 (Hereford)	42	33	1·080
Machynnlleth (Montgomery)	101	80	2·431
Newport (Pembroke)	78	67	0·834
Chancery (Cardigan)	73	63	0·735
Bromlys 1 (Brecon)	67	58	0·648
Pangbourne (Berks)	77	67	0·694
Bredenbury 2 (Hereford)	52	46	0·367
Haverfordwest (Pembroke)	82	76	0·228
Garreg 1 (Caernarvon)	89	93	0·088
Tenby (Pembroke)	67	70	0·066
Port Meirion (Merioneth)	73	81	0·416
Garreg 2 (Caernarvon)	71	80	0·536
Bromlys 2 (Brecon)	65	74	0·871
Ymstyllynn (Caernarvon)	40	51	1·330
Jeffreston (Pembroke)	61	89	5·227
Miscellaneous	4	10	2·571
17	1172	1130	24·933
Scott's data (Edinburgh)	56	44	1·960
Darwin's data (Kent)	40	39	0·013
v. Tschermak's data (Austria)	758	745	0·112
20	2026	1958	27·018

more plants of the same type were more frequent than they should have been on a basis of chance. A ratio of 1·5 for χ^2/n could be explained if, on an average, 50 % of seedlings reproduced themselves once vegetatively, just as the fluctuations in the sex ratios of human families would be greater if 50 % of all births were monozygotic twins. But I do not think the correction for clonal reproduction can have been more than 10 % except perhaps in the population at Newport, which actually did not give very divergent numbers.

It is certain that no obvious environmental effect exists on the thrum : pin ratio. The two most extreme populations were found on road banks within a few miles of one another. The situation is quite unlike that found in *Lythrum salicaria* (Haldane, 1936 *b*) where the frequencies of the three types in different localities were undoubtedly different. The reason is probably as follows. Suppose a single pin plant among a number of thrums. Then if there is an adequate opportunity for cross-fertilization its pollen will "take" on all the thrum plants, since it is probable that legitimate pollen tubes grow quicker than illegitimate (cf. Tseng, 1937). Thus equality will be almost if not quite restored in one generation. Whereas if there is only one long-styled *Lythrum* among a number of mid-styled and short-styled plants its pollen will only have twice the opportunities of the other types.

If a significant excess of thrum plants is ultimately found, the explanation is far from obvious. Darwin (1877, p. 36) found that when protected from flying insects (but not from thrips) pin plants set 19·2 seeds per capsule on an average as a result of fertilization, whilst thrums set only 6·2. If this were so we might expect an excess of pins in sparse primrose populations, such as furnished the "miscellaneous" group. But only further work will confirm or disprove this hypothesis.

SUMMARY

The ratio of thrum to pin plants among 2302 primroses did not differ significantly from equality. Individual populations did not diverge from equality to a significantly greater extent than could be expected as the result of sampling error.

REFERENCES

DARWIN, C. (1877). *The different forms of flowers on plants of the same species.* London.

GREGORY, R. R. (1915). "Note on the inheritance of heterostylism in *Primula acaulis* Jacq." *J. Genet.* **4**, 303.

HALDANE, J. B. S. (1936 *a*). "Linkage in *Primula sinensis*. A correction." *J. Genet.* **32**, 373.

—— (1936 *b*). "Some natural populations of *Lythrum salicaria*." *J. Genet.* **32**, 393.

—— (1938). "The approximate normalization of a class of frequency distributions." *Biometrika*, **29**, 392.

v. TSCHERMAK, E. (1923). "Ueber Varietäten und Specieshybriden bei *Primula*." *Proc. Int. Horticult. Congr. Amsterdam*, p. 139.

TSENG, HSIEN-PO (1937). "Pollen-tube competition in *Primula sinensis*." *J. Genet.* **35**, 289.

DE WINTON, D. & HALDANE, J. B. S. (1933). "The genetics of *Primula sinensis*. II. Segregation and interaction of factors in the diploid." *J. Genet.* **27**, 1.

(vi) Notes of Karl Pearson's Lectures on the Theory of Statistics, 1884–96*

BY G. U. YULE, F.R.S.

INTRODUCTION

In the following abstract of my notes on these early lectures the actual terminology has been in general retained: much of it, e.g. the terms "centroid" (centre of gravity) and "swing radius" (radius of gyration, root-mean-square radius), is conveyed direct from Professor Pearson's lectures to engineers, and might well puzzle a modern statistician. Sentences or paragraphs placed in quotation marks are direct quotations from the notes. These lectures are so closely related to the early memoirs that it is desirable to keep in mind the dates by which these were completed, as indicated by the dates of receipt by the Royal Society. The more important, for which detailed references are given below, are:

(1) Dissection of compound Normal Distribution	Oct. 18th, 1893
(2) Skew Variation	Dec. 12th, 1894
(3) Note on Regression and Inheritance in the case of two parents. *Proc. Roy. Soc.* LVIII, pp. 240–241	June 5th, 1895
(4) Regression, Heredity and Panmixia	Sept. 28th, 1895
(5) On the Probable Errors, etc. (Pearson and Filon)	Oct. 18th, 1897

The memoir on the dissection of a compound normal distribution had then been completed a year before the first course began; the memoir on skew variation was completed at the close of the first term of that course; the first note on correlation (including a partial regression equation) in which it is stated that ill health had delayed the completion of the full memoir, towards the end of the summer term and the full memoir itself, in which the "best" value of the correlation coefficient (i.e. the method of maximum likelihood value, obtained from the product-sum formula) was given for the first time, at the end of the

* The following article was very kindly prepared by Mr Yule as an additional Appendix to my memoir of Karl Pearson (*Biometrika*, **28**, 193–257 and **29**, 161–248). It will be reprinted with the rest of the memoir when this is published shortly as a separate volume by the Cambridge University Press. [E. S. P.]

following long vacation. The long and important memoir on probable errors was not finished till after the end of the second course. Dates only occur rather erratically in the notes: they have been given when they place the work in a given term.

The first course opened with a brief outline sketch of history, leading up to a "Kollektiv-mass" definition of statistics. Among the works bearing on theory to which we were referred those of Zeuner, Lexis, Edgeworth, Westergaard and Levasseur might be expected: but would any other lecturer have thought of suggesting the study of Marey's *La Méthode graphique dans les Sciences Expérimentales* (1878 and 1885)? Karl Pearson was an enthusiast for graphic representation and thought in graphic terms. After this introduction, theory proper was begun with Bayes' Theorem—not with the correlation approach of later years, which would hardly have been likely then. Thereafter we were taken to frequency distributions, means and moments in general, and a classification of theoretical forms was suggested. The binomial series followed, and the normal curve: for an area table of the normal curve based on the standard deviation reference could only be made to the short table printed in the Gresham Lecture Notes. The discussion of the error in the standard deviation caused by an error in any given ordinate, when the standard deviation is determined from the moment of any given order, I do not recall seeing given elsewhere. There was then a reversion to the moment problem and the moments of the binomial series: the correcting terms in these, which seem to have puzzled Professor Fisher,* are simply the correcting terms required to give the moments of the representation by histogram or by frequency polygon—i.e. the moments of the graphic figure—in terms of the moments of weighted ordinates. Some problems on standard deviations evidently concluded the work of the first term. The second term, apparently after completing the last subject, began with reference to the sources from which examples of skew distributions could be drawn: some of such distributions are probable compound, and this led to a series of notes on memoir(1). Some problems on inheritance were then interpolated, and after this followed the derivation of frequency curves from the binomial series and the hypergeometric series, in fact the work of memoir(2), which had been completed only in the previous December. No date in the notes indicates where the work of this term ends, but the notes are so extensive that the lectures must almost have continued into the summer term. In that term at least will have followed the work on correlation, not completely published till the end of the following September.

A straightforward, organized, logically developed course clearly could hardly then exist when the very elements of the subject were being developed: there are occasional breaches of continuity, or divergences to subsidiary or illustrative problems that were interesting the lecturer: or a difficult problem, e.g. the moments of the hypergeometric series, might be simply dropped and taken up again a little later. In the following year this feature becomes still more marked. Memory will not now recall exactly what happened, but the members of the class were probably largely engaged in practical work: this is the only way in which I can account for the lectures apparently not beginning till November 21st. It will be seen that such practical work evidently accompanied, or was interpolated in, the lecture course at a later stage to test the results arrived at in the lectures on skew correlation, which in conjunction with the practical work formed a piece of pure research. It will be noted also that some lectures on probable errors were inserted in January 1896 in the middle of those on skew correlation, probably while the test-work was going on.

The lectures on Theory of Error in May 1896 are of interest: the first set of experiments (bisection of line) on which the memoir(6) of 1902 was founded were carried out that summer (1896) ((6), p. 243). The curious result for a distribution compounded of two half normal curves I do not remember seeing elsewhere. One other point calls for a late apology from me: when writing the note "On Reading a Scale" (*Jour. Roy. Statist. Soc.* 1927) I had no recollection that Karl Pearson had directed my attention to preferences and avoidances of particular digits thirty-one years before! The note came as a complete surprise.

Sheppard's Theorem, which concludes the notes, must presumably have been personally communicated to Pearson, as it was not published till some two years later.

* See footnote to an article on W. F. Sheppard, *Annals of Eugenics*, VIII (1937), pp. 9–10.

SUMMARY OF THE LECTURES

The material is taken from G.U.Y.'s notes of that date, now preserved in the Department of Statistics at University College

Session 1894–1895

Original meaning of word "statistics". Outline history: Graunt, Petty, de Witt, Breslau mortality statistics, Halley, Kersseboom, Déparcieux, Süssmilch, Achenwall, Playfair, Laplace, Quetelet; Edgeworth, Galton and Weldon; Mayr, Block, Gabaglio. Definition: "Statistics is simply a name used for aggregate measurements of any facts whatever, whether social, physical or biological. The theory of pure statistics is that branch of mathematics which deals with the compilation, representation and handling of numerical aggregates—and this independently of the facts which the numbers represent. Applied Statistics is the application of the methods of pure statistics to special classes of facts—biological, physical or political observations for example." Works on theory cited: Zeuner, Lexis, Edgeworth, Westergaard, Levasseur, Marey.

Bayes' Theorem: the fundamental principles assumed (1) permanence of statistical ratios, (2) equal distribution of ignorance (Note: "At the Gresham Lectures the audience were asked to guess how many white balls there were in a bag of 20 black and white: the guesses grouped round 10, quite unreasonably.") Examples of Bayes' Theorem. "The statistically supported principle of the equal distribution of ignorance."

Frequency curves: continuous and discontinuous distributions: great variety of forms. Mean, median and most frequent value or mode. Deviations, different meanings. Quartiles, percentiles, Galton's Ogive: disadvantage of representation by percentiles.

Moments: mean error, mean pth deviation: "probable deviation" in excess and defect, "probable error" in this sense. The standard deviation, defined as "the swing radius of the curve about the centroid vertical." Modulus. Skewness, measured by (mean − mode)/standard deviation.

Forms of frequency curve classified in five types: 1. Mode at one end of base. 2. Curve rising at a definite angle to base, range limited or unlimited at other end. 3. Range limited in one direction, but curve starting tangential to base. 4. Skew, range unlimited in both directions. 5. Symmetrical, range limited or unlimited. A function is wanted to cover all these forms. Brief reference to frequency surfaces or correlation surfaces for two or more variables.

The binomial distribution: experiments show that "the mathematically *possible* distribution is the experimentally *probable* distribution." Illustrations. Representation by polygon, becoming a curve in the limit when n is large. Binomial machine.

Normal curve: s.d. may be determined either from (1) the co-ordinates of the centroid, i.e. of the centre of gravity of the area between the curve and its base line, or (2) from the swing radius about the centroid vertical: it is usual to take the areas of the elementary trapezia as concentrated on their mid-ordinates, but corrections will be required. Moments of the normal curve. Error in the s.d. caused by an error in any given ordinate, when the s.d. is determined from a moment of any given order. Area table of normal curve (reference to Gresham Lecture Notes) and its use: use of three times the s.d. as limit for likely deviations. Fitting from mean deviation and from quartiles. Rough test of "goodness of fit" by ratio Σ (errors of fit without regard to sign)/N: values for 12 actual distributions, ranging from about 6 to 13·5 per cent.

The standard deviation of the standard deviation for a normal distribution.

"Reduction of the moments of a curve treated as a series of trapezia to its moments when the elementary areas are concentrated along ordinates": (this is the heading in my notes, but the problem taken is to express the moments of the representation by histogram (rectangles), or by frequency polygon, in terms of the moments of weighted ordinates: the work is that of the memoir(2), pp. 348 *et seq.*).

Moments of the binomial series. Complete fitting of a binomial, taking the interval c between ordinates as unknown, as well as n, p and q.

Determination of the standard deviation of a ratio z_1/z_2 in terms of the s.d.'s of z_1 and z_2, when deviations are assumed small compared with the means and z_1, z_2 are independent. (Dec. 20th.)

General result for any function of the z's.

Statistical sources of skew distributions: homogeneous distributions and compound curves.

The dissection of a compound normal distribution: notes on the memoir(1).

Some problems in inheritance for a population following the normal law. (1) Parents of deviation x in a population with s.d. σ_1 give rise to a fraternity with mean x/n and s.d. σ': what is the distribution of the next generation? Generalization for successive generations. Biological deductions. (2) A normal sub-population of parents is selected with mean h and s.d. Σ: what is the distribution of the offspring?

The slope-relation between the normal curve and the symmetrical binomial. The slope relation for general binomial: the resulting skew curve: its moments and method of fitting (memoir(2)). The empirical (one-third) relation between mean, median and mode. Reduction of this distribution to normal curve. Edgeworth's distribution (generalized normal curve).

Generalization of binomial by removing the assumption that "contributory causes" are not independent: "the theory of interdependence will be based on the assumption that the independence of contributory causes is limited by a limited material from which to produce effect," e.g. drawing r balls from a bag containing pn black, qn white. The (hyper-geometric) series in this case: moments, slope relation, and resulting curves (memoir(2)).

Correlation: notion of x, y, z being correlated from each being a function of $\rho_1 \rho_2 \ldots \rho_n$: assuming that (1) all variations in ρ's are small, (2) follow the normal law, (3) are independent, the general expression for the normal correlation distribution is deduced.

Special case of two correlated variables: expression of the parameters in terms of N, σ_1, σ_2, and r, "Galton's function". Properties of the distribution; regressions, s.d.'s of arrays. The "best" value to give r, deduction of the product-sum formula. "This method of reckoning r has not been used for any system of correlated organs, but approximate methods, by no means the best, have been used by Galton, Weldon and Edgeworth."

The standard deviation of the coefficient of correlation for a normal distribution (the erroneous value, in effect the standard error for determinate values of the standard deviations, corrected in memor(5), p. 242).

Contour lines of normal surface: Galton's determination of the contours as ellipses and estimation of r from the vertical tangents. The slope of the principal axes: estimation of r from these axes, determined say by cutting the ellipses by circles. Estimation of r from the s.d. of arrays. The s.d.'s in direction of principal axes: expression for the normal surface referred to the principal axes. The property that the proportion of frequency falling outside the ellipse χ is $e^{-\frac{1}{2}\chi^2}$ (Bravais): "probable ellipse" and "standard ellipse." The proportion of frequency lying within a circle of given radius round the mode: table: approximate formulae.

Normal distribution for three variables: deduction of the general expression in terms of standard deviations and correlations. Correlation between father, mother and offspring as an example. The regression equation. The three-variable surface referred to principal axes: the contour ellipsoid: proportion of frequency outside a given ellipsoid: short table. The chance that an observation lies in a particular cone or polar element spreading out from the centre.

Session 1895–1896

"The following notes on skew correlation were begun Nov. 21st, 1895." It is not clear why they begin so late in term. "Up to the present no theory of skew correlation exists and, although numerous observations involving the frequency of two variables are easily seen at once to be skew, no correlation surface has yet been fitted to such distributions. Hence whatever theory we adopt must be regarded as a trial, and its only justification must be that it suffices to describe observed statistics." Three different approaches were tried.

I. Hypothesis that the variations in two directions at right angles are independent, suggested by the normal surface. Relations of moments and product moments: the directions of independent variation are the principal axes. Problem: "Both independent variations being of the same type, what must that type be in order that every vertical section of the surface shall be also of the same type?": proof that it must be the normal law. First four moments of such a surface about the principal axes in terms of moments parallel to the axes of measurement. Order of work: (1) Find first four moments of total distributions. (2) Determine principal axes. (3) Convert moments calculated to moments about principal axes, which determine the distributions for principal axes, say $\phi(x)$, $f(y)$. (4) $z=\phi(x)f(y)$ is the equation to the surface. Note added at end: "The previous assumption (independent variation in two directions at right angles) was found not to work in the case of Perozzo's age-at-marriage surface. This led to trial of a more extensive assumption."

II. (Feb. 1st, 1896.) Two directions of independent variation, not necessarily orthogonal. General case: deduction of condition that for n variables there shall be n directions of independent variation, not necessarily orthogonal. Special case of two variables: the directions of independent variation must be conjugate diameters of the ellipse of inertia: moments and product moments and their relations. Concluding note dated March 1896: "This theory was tried on a surface correlating the heights of barometers at two different stations and failed."

III. Hypergeometrical surface. A bag contains n balls, pn white, qn black: m balls are drawn (without return) and then a second lot of m' balls. What is the chance that r of the m and s of the m' are black? "One of the advantages of this form of the correlation ordinate is that the surface summed in the direction of either axis of correlation gives the very expression from which we have deduced skew variation curves; in other words we shall expect the curve formed by the sums parallel to either axis to be the skew curves we have already found to be applicable. Moreover, any section parallel to either of the axes of correlation is also a hypergeometrical series, i.e. a close approximation to a skew curve." Attempts were made on two different lines to deduce a curved surface (1) by means of a slope-relation as for skew curves, (2) by approximating by Stirling's theorem. The results are summarized as follows: "The attempt to get a surface parallel to the polyhedron of correlation which arises in ordinary chance problems leads us to values of the differentials of z (the ordinate) which, as far as we see, cannot be integrated. But these values of z show us two points of interest (1) that there is only one direction in any skew correlation surface in which the line of modes is a straight line and (2) in any other direction it is a cubic curve. Approximating to the ordinate of the same surface by Stirling's theorem we obtain an equation which confirms the results of our first two trials (i.e. I and II), for there is no possibility of breaking up the expression into factors. The fact that the curve of regression is a cubic is also confirmed, and the form that may approximately be given to it, at least in the neighbourhood of the mode is also confirmed." (March 26th, 1896.)

Reproductive Selection (April 23rd, 1896). The deduction of the formulae (i) to (iv) published in the "Note on Reproductive Selection", *Proc Roy. Soc.* LIX, p. 301, received Feb. 13th, 1896.

The following section on probable errors is dated at end January 1896· this suggests the lectures were interpolated while the practical work on skew correlation was being done:

Probable errors of skew curve constants: with a note on the differentials of Gamma-functions, and a short table. General theorem on the probable error of a mean: the approach is that of the method used in the memoir by Pearson and Filon(5). This is followed by a similar General Theorem on the Probable Error of any Constant.

Theory of Errors (May 14th, 1896). Classification of types of error: theoretical errors, instrumental errors, personal errors. "Astronomers do not appear to have ever dealt with personal equations by the experimental method." There are two points, the deviation of an observer from the truth and the mean deviation of his observations from his own mean. "Neither of these points has been really looked into. Error of judgment and variability

of judgment are both important." Preferences for particular digits noted: "for example, in 1000 readings by the same observer 0·3 occurred only 30, but 0·4, 170 times." Accidental or irregular errors: the problem of the rejection of observations. Considerations arising as to the assumption of normality: the criteria never exactly fulfilled. To test effect of slight divergence from the normal, a distribution is considered composed of two half normal curves, numbers of observations above and below mean n_1 and n_2, s.d.'s σ_1 and σ_2. σ_2 is then written $\sigma_1 \times \alpha$. α is assumed small and $n_1 - n_2$ also small, and the moments evaluated, with the final result

$$\beta_2 - 3 = 3\frac{\pi}{2}\beta_1.$$

Probable errors of α and of the criterion in this neighbourhood.

Sheppard's Theorem for the correlation in terms of the frequencies in the four quadrants of a normal distribution divided at the medians; geometrical proof (*Phil. Trans. Roy. Soc.* A, CXCII (1898)).

REFERENCES

(1) "Contributions to the Mathematical Theory of Evolution." *Phil. Trans. Roy. Soc.* A, CLXXXV (1894), 71–110. [This is the first of a series of memoirs, afterwards numbered II, III, etc., entitled "Mathematical Contributions to the Theory of Evolution."]

(2) II. "Skew Variation in Homogeneous Material." *Phil. Trans. Roy. Soc.* A, CLXXXVI (1895), 343–414.

(3) "Note on Regression and Inheritance in the case of Two Parents." *Proc. Roy. Soc.* LVIII (1895), 240–41.

(4) III. "Regression, Heredity and Panmixia." *Phil. Trans. Roy. Soc.* A, CLXXXVII (1896), 253–317.

(5) IV. "On the Probable Errors of Frequency Constants and on the Influence of Random Selection on Variation and Correlation." *Phil. Trans. Roy. Soc.* A, CXCI (1898), 229–311.

(6) "On the Mathematical Theory of Errors of Judgment, with Special Reference to the Personal Equation." *Phil. Trans. Roy. Soc.* A, CXCVIII (1902), 235–99.

(vii) Frequency Curves and Correlation. By W. PALIN ELDERTON. Third Edition. Cambridge, at the University Press, 1938. Price 12*s.* 6*d.*

In the following review only those parts of the book will be dealt with, which have been altered since the former edition. For that reason we are specially interested in chapters 10, 11 and 12, which, according to the preface of this third edition, have in many respects been rewritten. The headings of these chapters are: "Standard errors", "The test of goodness of fit" and "The correlation ratio–contingency". In the chapter concerned with the test of goodness of fit, R. A. Fisher's opinion about the χ^2-test is explained in greater detail than in earlier editions. The author has the sound opinion that: "when we merely want to compare several graduations of the same distribution we can often stop our work after the calculation of χ^2." The methods for deducing standard errors (chapters 10 and 12) are more exact than before and treated in greater detail.

In the other chapters, which have not been rewritten but only altered in one or another respect, we observe a short historical note about the normal curve of error, being a transition type of the Pearson curves. Further, in chapter 6, reference is made to the underlying theory of the A-series, which as in the former edition is stated as an alternative to the Pearson curves. In chapter 3 the method for working out the moments by iterated summations is simplified in the well-known way by first computing the factorial moments. A new and valuable Appendix (number 5 in the new edition) has been added, containing a short description of other methods than that of moments for estimating unknown constants. The methods described are (1) that of least squares, (2) that of maximum likelihood and

(3) the minimum χ^2-method. It is also worth mentioning that in Appendix 2 some account is given not only of the complete "beta"- and "gamma"-functions but also of the incomplete ones, and references are made to tables of these functions. At the end of the book there is, as in the former edition, a table of log $\Gamma(p)$ but the new edition also contains brief tables of the normal curve of error and of the χ^2-distribution.

The new edition, as the earlier ones, is mainly a textbook for computers and specially for those wishing to apply Pearson curves to empirical distributions. Much new beyond that contained in the former edition has not been added to these technical sides, and the disposition of the book is maintained. For the further statistical analysis the author has in the new edition made valuable additions and alterations, the most important of which have been mentioned above.

O. LUNDBERG.

Stockholm, 1938.

Vol. XXX. Parts III and IV January, 1939

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF BIOLOGICAL PROBLEMS

FOUNDED BY

W. F. R. WELDON, FRANCIS GALTON AND KARL PEARSON

EDITED BY

EGON S. PEARSON

IN CONSULTATION WITH

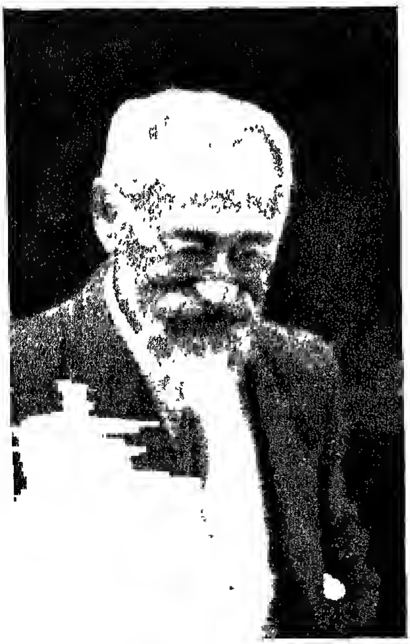
HARALD CRAMÉR J. B. S. HALDANE
MAJOR GREENWOOD G. M. MORANT
JOHN WISHART

ISSUED BY THE BIOMETRIKA OFFICE
UNIVERSITY COLLEGE, LONDON
AND PRINTED AT THE
UNIVERSITY PRESS, CAMBRIDGE

PRINTED IN GREAT BRITAIN

[Issued 25 January 1939] *Reprinted by offset-litho 1952*

On Dartmoor during a fishing holiday, April 1936.



On the bowling green, May 1937.

William Sealy Gosset

VOLUME XXX JANUARY, 1939 PARTS III AND IV

WILLIAM SEALY GOSSET, 1876–1937

The two appreciations which follow have been written from somewhat different angles. The first is by a younger colleague and friend at the St James' Gate Brewery, who for a number of years was in close contact with Gosset in Dublin, both in and out of the brewery. The friendship of the second writer is one which grew through a correspondence that roved at length over statistical methods and theories. If in some places the articles overlap, this will only help to emphasize certain events or characteristics which independently we have felt impelled to record.

Both of us would like to express our warmest thanks to the many friends who have helped us, and in particular to Mrs W. S. Gosset and Mr E. Somerfield.

L. McM., E. S. P.

(1) "STUDENT" AS A MAN

BY L. McMULLEN

WILLIAM SEALY GOSSET was the eldest son of Colonel Frederic Gosset, R.E., and was born at Canterbury in 1876. In 1906 he married Marjory Surtees Phillpotts, daughter of the late headmaster of Bedford School, and they had one son and two daughters. He died on 16 October 1937, and was survived by both his parents, his wife and children, and one grandson.

He was educated at Winchester, where he was a scholar, and New College, Oxford, where he studied chemistry and mathematics.

He entered the service of Messrs Guinness as a brewer in 1899.

It is not known exactly how or when "Student's" interest in statistics was first aroused, but at this period scientific methods and laboratory determinations were beginning to be seriously applied to brewing, and it is obvious that some knowledge of error functions would be necessary. A number of university men with science degrees had been taken on, and it is probable that "Student", who was the most mathematical of them, was appealed to by the others with various questions and so began to study the subject. It is known that he could calculate a probable error in 1903. The circumstances of brewing work, with its variable materials and susceptibility to temperature change and necessarily short series of experiments, are all such as to show up most rapidly the limitations of large sample theory and emphasize the necessity for a correct method of treating

small samples. It was thus no accident, but the circumstances of his work, that directed "Student's" attention to this problem, and so led to his discovery of the distribution of the sample standard deviation, which gave rise to what in its modern form is known as the t-test. For a long time after its discovery and publication the use of this test hardly spread outside Guinness's brewery, where it has been very extensively used ever since. In the Biometric school at University College the problems investigated were almost all concerned with much larger samples than those in which "studentizing", as it was sometimes called, made any difference. Nevertheless, although their lines of research diverged somewhat rapidly, the close statistical contact and personal friendship between Karl Pearson and "Student", which began during his year at University College, were only terminated by death.

The purpose of this note is not however to give an account of "Student's" statistical work, but to try to give a more general impression of the man himself. Although his public reputation was entirely as a statistician, and he was acknowledged to be one of the leading investigators in that subject, his time was never wholly and rarely even mainly occupied with statistical matters. For one who saw enough of him to know roughly how his time was spent both at work and at home, it was very difficult to understand how he managed to get so much activity into the day. At work he got through an enormous amount of the ordinary routine of the brewery, as well as his statistics. Until 1922 he had no regular statistical assistant, and did all the statistics and most of the arithmetic himself; later there was a definite department, of which he was in charge till 1934, but throughout he did a great deal of arithmetic and spade-work himself. It might be supposed from the amount he did in the time that he was unusually good at arithmetic and the arrangement of work; such, however, was not the case, for his arithmetic frequently contained minor errors. In one of his obituary notices a tendency to do work on the backs of envelopes in trains was mentioned, but this tendency was not confined to trains; even in his office much work was done on random scraps of paper. He also had a great dislike of the tabulation of results, and preferred to do everything from first principles whenever possible. This preference led in certain instances to waste of time in routine work, but was of assistance in maintaining that flexibility and speed of attack on new problems which was so characteristic of him. An actual example would need too much explanation of relevant circumstances, but I can vouch for the analogical truth of the following. If a body performs simple harmonic motion with acceleration μ per unit displacement, it may readily be shown that the period of a complete oscillation is $2\pi/\sqrt{\mu}$. Hence, in the case of a simple pendulum $t = 2\pi\sqrt{(l/g)}$ and $l = gt^2/4\pi^2$, where l is the length of the pendulum and g the acceleration due to gravity. If it were necessary to calculate the lengths of pendulum corresponding to different periods as a routine matter, most people would evaluate $g/4\pi^2$ for their locality and always multiply t^2 by

this numerical constant, which would be about 24·85. "Student" would probably have started from $2\pi/\sqrt{\mu}$ every time. If therefore he had suddenly wanted to calculate the period of oscillation of a weight on a stretched spring he could have done it, whereas the man who only remembered that $l = 24{\cdot}85t^2$ for a pendulum would be unable to tackle the problem without much more preliminary work.

His method was, of course, not necessarily the most suitable for others not aspiring to the same degree of versatility. Perhaps it is not altogether fanciful to compare the two methods with the organic evolution of, say, the human hand, the most versatile object known, and the construction of some highly efficient but absolutely specialized piece of machinery. I do not mean to imply that he gave this explanation, or was even altogether conscious of it. When he handed over to me a routine calculation which he had done for many years, I was astonished to find that he had written out every week an almost unvarying form of words with different figures. To my question, "Why ever don't you get a printed form?" he did not reply, "Doing it from first principles every time preserves mental flexibility". He would have considered such a remark unbearably pompous. He said, "Because I'm too lazy", to which I replied, "Well, I'm too lazy not to."

To many in the statistical world "Student" was regarded as a statistical adviser to Guinness's brewery; to others he appeared to be a brewer devoting his spare time to statistics. I have tried to show that though there is some truth in both of these ideas they miss the central point, which was the intimate connexion between his statistical research and the practical problems on which he was engaged. I can imagine that many think it wasteful that a man of his undoubted genius should have been engaged in industry, yet I am sure that it is just that association with immediate practical problems which gives "Student's" work its unique character and importance relative to its small volume. On at least one occasion he was offered an academic appointment, but it is almost certain that he would not have been a successful lecturer, though perhaps a good individual teacher; nor is it likely that his research work would have flourished in more academic circumstances; his mind worked in a different way.

The work in connexion with barley breeding carried out by the Department of Agriculture in Ireland, in which Messrs Guinness took a prominent part, enabled "Student" to get that first-hand experience of yield trials and agricultural experiments generally which contributed so largely to his great knowledge of the subject. He did not merely sit in his office and calculate the results, but discussed all the details and difficulties with the Department officials, and went round all the experiments before harvest, when a "grand tour" is annually carried out by the Department, the brewery, and sometimes statisticians or others interested from England or abroad. As well as the work carried out at the actual cereal station near Cork, three or four varieties of barley are grown in

$\frac{3}{4}$ or 1 acre plots at ten farms representing all the principal barley-growing districts of Ireland, so a visit to all of them entails a fairly comprehensive inspection of the crops.

"Student" took a great deal of interest in this work from the beginning and correspondence shows that he discussed the results of these tests with Karl Pearson at great length when he went to study with him at University College in 1906.

In the last ten years or so of his time in Ireland he played a leading part in these investigations, and thus had a perhaps unique opportunity of following experimental varieties from sowing through growing and harvest to malting and brewing results, and also of carrying out or supervising all the relevant mathematical work. At one time he also made some barley crosses in his own garden, and accelerated their multiplication by having one generation grown in New Zealand during our winter. These crosses were known as Student I and II, and have now been discarded as failures, the inevitable fate of the large majority. With characteristic self-effacement he was the first to point out that they were not worth going on with.

He also made frequent visits to Dr E. S. Beaven, whose work on barley breeding is well known, and discussed every aspect of yield trials with him. These visits were undoubtedly very useful, and although Dr Beaven is never tired of protesting that he is no mathematician and does not understand "magic squares" or "birds of freedom", which he prefers to the more orthodox expressions, he has a vast experience of agricultural trials and is very quick to see the weak point of any experiment.

In spite of the quantity of work "Student" did he was never in a hurry or fussed; this was largely due to the absence of lag when he turned his mind to a new subject; unfortunately others were not always equal to this. He would ring one up on the phone and plunge straight into some subject which might have been discussed some days previously. The slower-witted listener would probably lose the thread of his discourse before realizing what it was about and would ignominiously have to ask him to begin again. I have many times seen him hard at it on a Monday morning, but at first meeting it was always "How did the sailing go?" "Well, did you catch any fish?", and he would recount any notable event of his own week-end before plunging into the very middle of some subject. I never heard him say "I'm busy".

"Student" had many correspondents, mostly agricultural and other experimenters, in different parts of the world. He took immense pains with these and often explained points to them at great length when he could easily have given a reference. His letters contain some of his clearest writing, and the more difficult points are often better elucidated than in his published papers.

Karl Pearson emphasized the fact that a statistician must advise others on their own subject, and so may incur the accusation of butting in without

adequate knowledge. "Student" was particularly expert at avoiding any such disagreement; usually he was such an enthusiastic learner of the other's subject that the fact that he was giving advice escaped notice.

The reader will by now have realized that "Student" did a very large quantity of ordinary routine as well as his statistical work in the brewery, and all that in addition to consultative statistical work and to preparing his various published papers. It might thus be thought that he could have done nothing else but eat and sleep when at home; this, however, was far from being the case, and he had a great many domestic and sporting interests. He was a keen fruit-grower and specialized in pears. He was also a good carpenter, and built a number of boats; the last, which was completed in 1932, and on whose maiden voyage I had the honour to be nearly frozen to death, was equipped with a rudder at each end by means of which the direction and speed of drift could be adjusted—an advantage which will be readily appreciated by fly-fishermen. This boat with its arrangement of rudders was described in the *Field* of 28 March 1936. In his carpentry he showed preferences analogous to his mathematical ones previously mentioned; he disliked complicated or specific tools, and liked to do anything possible with a pen-knife. On one occasion, seeing him countersinking screw-holes with a pocket-knife, I offered him a proper countersink bit which I had with me, but he declined it with some embarrassment, as he would not have liked to explain or perhaps could not have explained why he preferred using the pen-knife. Out of doors he was an energetic walker and also cycled extensively in the pre-war period. He did a lot of sailing and fishing. For his last boat he had a most unconventional sail, which cannot be exactly described under any of the usual categories; it was illustrated in the *Field* article referred to above.

In fishing he was an efficient performer; he used to hold that only the size and general lightness or darkness of a fly were important; the blue wings, red tails and so on being only to attract the fisherman to the shop. This view was more revolutionary when I first heard it than it is now. He was a sound though not spectacular shot, and was well above the average on skates. Until the accident to his leg in 1934 he was quite a regular golfer, and once went round a fairly difficult course in 85 strokes and 1¼ hours by himself. He used a remarkable collection of old clubs dating at least from the beginning of the century. In the last few years since his accident he took up bowls with great keenness, and induced many other people to play as well. One of his last visits to Ireland was with a team which he had organized at the new brewery at Park Royal.

On top of all this he knew as much as most people of the affairs of the world in general and of what was going on about him. It became very difficult to imagine how he found 24 hours in any way a sufficient length for the day. His wife certainly organized things so that the minimum amount of time was wasted, but even so few people could approach such activity in quantity or diversity.

In personal relationships he was very kindly and tolerant and absolutely

devoid of malice. He rarely spoke about personal matters but when he did his opinion was well worth listening to and not in the least superficial.

In the summer of 1934 he had a motor accident and broke the neck of his femur. He had to lie up for three months, of course working at statistics, and was a semi-cripple for a year. This was particularly irksome for such an active man, as was the sheer unnecessariness of the accident, for he ran into a lamp-post on a straight road, through looking down to adjust some stuff he was carrying; but with great hard work and persistence he eventually reduced the disability to a slight limp.

At the end of 1935 he left Ireland to take charge of the new Guinness brewery in London, and I saw comparatively little of him after that. The departure from Ireland of "Student" and his family was a great loss to many who had experienced their hospitality.

His work in London was necessarily very hard and accompanied by all the vexations inevitably associated with a big undertaking in its first stages, before any settled routine has been established; nevertheless, he still found time to continue his statistical work and wrote several papers.

His death at the comparatively early age of 61 was not only a heavy blow to his family and friends, but a great loss to statistics, as his mind retained its full vigour, and he would undoubtedly have continued to work for many more years.

I am painfully conscious of the inadequacy of this sketch, which cannot hope to convey more than a faint impression of his unique personal quality to those who did not know him, but it will have served its purpose if it helps any readers to grasp the essential unity and directness of the personality which lay behind such widely varied manifestations.

(2) "STUDENT" AS STATISTICIAN

By E. S. PEARSON

For many years after the publication of his first paper in *Biometrika*, in 1907, the name of "Student" was associated in statistical circles with an atmosphere of romance. Those who knew him only through his written contributions must often have wondered who was this unassuming man, content to remain anonymous, who wrote so clearly and simply on so wide a range of fundamental topics. To those of us who came into touch with him personally, the knowledge that "Student" was W. S. Gosset did not altogether dispel that romantic impression. Here, in London, he would pay us visits from time to time at the old Biometric Laboratory on his way to Euston station to catch the Irish mail;

he would be wearing the grey flannel trousers that were a tradition of his Wykehamist schooldays and carrying a rucksack on his back. And then after a short hour's talk, perhaps on statistical subjects, perhaps on his garden experiments in cross-breeding, he would be off again to that wild Ireland where, in the "bad times", we had heard that gunmen were to be found hiding behind his hedges or even searching his house for arms. We had heard too of great exploits by members of his family of an entirely non-statistical character, of their boat-building and of their construction of a pair of water-skis which they used for walking over Kingstown Harbour.

My one short winter visit to Gosset's house at Blackrock, a few miles outside Dublin, would hardly by itself have cleared away this element of myth or made me appreciate fully the sterling values that lay beneath that friendly and unassuming exterior. We talked very little about statistics during my stay, and the strongest impressions remaining are of a morning spent among the immense vats and varied smells of the brewery; of drives out of town on misty evenings through the badly lit Dublin suburbs in that old, high two-seater Model-T Ford of his, christened "The Flying Bedstead"; of the warm hospitality of his fellow-brewers; and of a Saturday in the snow-covered Wicklow Mountains when, letting his folk go off to test the more exciting slopes, he patiently tried to teach me to ski on a short stretch of mountain road.

My real understanding of Gosset as a statistician began, as no doubt for many others, when I joined that wide circle of his scientific correspondents. Perhaps to the majority of these he has stood as the friend who, with a greater mathematical knowledge, helped them to understand the statistical approach to experimental problems. In my own case the position was a little different, as his endeavour was always to temper my mathematical reasoning with sane common sense. I can think of no other statistician who would have shown that interest and forbearance over many years to a young man who was continually posting to him the results of half-finished investigations for comment and criticism. In looking back through this correspondence I realize more clearly now than I could ever have done at the time what its value to me has been, and I can see how many of his ideas scattered through these letters have since almost unconsciously become part of my own outlook. I think this must be true also in the case of other persons with whom he corresponded, so that one can say that the last thirty years' progress in the theory and practice of mathematical statistics owes far more to "Student" than could be realized by a mere study of his published papers.

One of the striking characteristics of these papers, also of course evident in correspondence, was the simplicity of the statistical technique he used. The mean, the standard deviation and the correlation coefficient were his chief tools; hardly adequate for treating specialized problems it might be thought; yet how extremely effective in fact in his skilled hands! There is one very simple and

illuminating theme which will be found to run as a keynote through much of his work, and may be expressed in the two formulae:*

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\rho\sigma_x\sigma_y, \qquad \text{......(1)}$$

$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y - 2\rho\sigma_x\sigma_y. \qquad \text{......(2)}$$

Perhaps we may count as one of his big achievements the demonstration in many fields of the meaning of that short equation (2); as he wrote in 1923 (**11**, p. 273, but with modified notation):

> The art of designing all experiments lies even more in arranging matters so that ρ is as large as possible than in reducing σ^2_x and σ^2_y.

It is a simple idea, certainly, but I cannot doubt that its emphasis and amplification helped to open the way to all the modern developments of analysis of variance, and there may be some who have felt that where this technique runs a risk of defeating its ends by over-elaboration is just where that simple maxim has been set on one side. Recently I came across a short passage in a letter to a friend in Australia which refers to this theme and illustrates Gosset's own humorously modest outlook on his own contributions. He had just received a good deal of criticism of a paper he read in March 1936 before the Industrial and Agricultural Research Section of the Royal Statistical Society (**21**), particularly because of his advocacy of the half-drill strip method of agricultural experiment. This is essentially a method of comparison whose efficiency depends on maximizing correlation by taking the difference between the yields of neighbouring strips of the two varieties or treatments compared. He wrote:

> Meanwhile I...enclose the rough proof of what I said at the Statistical. You will gather from that that I am not in the fashion....Some years ago an American referred to difference treatment as "Student's" method and, though at the time I referred it to Noah, I am beginning to think that there is something in the name.†

Another point which must be borne in mind in gaining a real understanding of Gosset's character and outlook is that all his most important statistical work was undertaken in order to throw light on problems which arose in the analysis of data connected in some way with the brewery. The subject of statistics was in no sense a whole-time job for him, nor, on the other hand, was it his hobby as it might perhaps be described in the case of W. F. Sheppard; he undertook theoretical investigations only when he or his colleagues were faced with difficulties which needed solution along statistical lines. Rapid if less accurate methods appealed to him because in much heavy routine work it was a question of finding such methods or of making no attempt at statistical treatment. He was in no hurry to see his results in print, and several of his papers in *Biometrika* were written in response to an editorial request rather than on his own initiative. In two cases at least, which I shall refer to below, he was using methods in the brewery ten years before publication was undertaken. He was indeed the ideal

* σ_x, σ_y, σ_{x+y} and σ_{x-y} are the standard deviations of x, of y, of $x+y$ and of $x-y$ respectively, and ρ is the coefficient of correlation between x and y.

† See (**14**, p. 709).

servant of his firm, and part of the value of his life's work would need to be recorded in a history of progress gained by scientific research in industry rather than in the pages of *Biometrika*.

Yet in spite of the fact that only a small part of his time was taken up with statistics, Gosset had a wonderful power of "getting there first" before the more professional statisticians. Perhaps it was because his greater detachment meant a continual freshness of mind. It is this characteristic, as well as those others I have mentioned, that I shall try to bring out in my description of his work in the following pages.

EARLY STATISTICAL INVESTIGATIONS

Gosset became one of the brewers of Messrs Arthur Guinness Son and Co., Ltd., in 1899. The firm had shortly before initiated the policy of appointing to their staff scientists trained either at Oxford or Cambridge, and these young men found before them an almost unexplored field lying open to investigation. A great mass of data was available or could easily be collected which would throw light on the relations, hitherto undetermined or only guessed at in an empirical way, between the quality of the raw materials of beer, such as barley and hops, the conditions of production and the quality of the finished article. With keen minds playing round the interpretation of these data it was almost inevitable that before long the need was realized of some understanding of the theory of errors. No doubt during the first few years of his appointment Gosset was mainly occupied with learning the routine work of his job, but once this knowledge had been gained it was natural that he, as the most mathematical of the younger brewers, should give his attention to the question of error theory. He seems to have made use of the following books: G. B. Airy, *Theory of Errors of Observations*; Lupton, *Notes on Observations*; M. Merriman, *The Method of Least Squares*.

By 1904 he had made himself sufficiently familiar with the subject to draw up a *Report* on "The Application of the 'Law of Error' to the work of the Brewery". This document, dated 3 November 1904,* opens with some paragraphs which set out in simple terms a case for the introduction of statistical method in large-scale industry. They are worth quoting since they might be put before many a board of directors to-day with just as much cogency as they were put 34 years ago in Dublin:

> The following report has been made in response to an increasing necessity to set an exact value on the results of our experiments, many of which lead to results which are probable but not certain. It is hoped that what follows may do something to help us in estimating the Degree of Probability of many of our results, and enable us to form a judgment of the number and nature of the fresh experiments necessary to establish or disprove various hypotheses which we are now entertaining.

* I am extremely grateful to the firm for giving me permission to see and quote from this and other records available in their Dublin brewery.

When a quantity is measured with all possible precision many times in succession, the figures expressing the results do not absolutely agree, and even when the average of results, which differ but little, is taken, we have no means of knowing that we have obtained an actually true result, and the limits of our powers are that we can place greater odds in our favour that the results obtained do not differ more than a certain amount from the truth.

Results are only valuable when the amount by which they probably differ from the truth is so small as to be insignificant for the purposes of the experiment. What the odds should be depends:

(1) On the degree of accuracy which the nature of the experiment allows, and

(2) On the importance of the issues at stake.

It may seem strange that reasoning of this nature has not been more widely made use of, but this is due:

(1) To the popular dread of mathematical reasoning.

(2) To the fact that most methods employed in a Laboratory are capable of such refinement that the results are well within the accuracy required.

Unfortunately, when working on the large scale, the interests are so great that more accuracy is required, and, in our particular case, the methods are not always capable of refinement. Hence the necessity of taking a number of inexact determinations and of calculating probabilities.

The *Report* then introduces the error curve and discusses some of its properties. The curve is written in Airy's form

$$y = \frac{1}{c\sqrt{\pi}}\, e^{-x^2/c^2}, \qquad \ldots\ldots(3)$$

where c is the modulus. The method is given for estimating c from a sample of n observations, by calculating (*a*) the mean deviation, (*b*) the mean square deviation (dividing by $n-1$), and using the appropriate correcting factors. It is stated that (*b*) gives a better value "in proportion 114/100".* A numerical example is given and it is suggested that both methods (*a*) and (*b*) should be used to check one another. There is next some discussion given to what was then clearly a most important practical problem in the brewery: the size of sample needed to make the odds that the mean lay within desired limits sufficiently large. Chauvenet's criterion for the rejection of extreme observations is quoted, as well as the modulus of the estimate of c (obtained by the mean square process), namely $c/\sqrt{(2n)}$.

All this is simply Airy or Merriman put by Gosset into the form most useful for his fellow brewers. What, however, shows a flash of his own insight is the use which he makes of Airy's theorems on the "Error of the result of the addition (or subtraction) of fallible measures". Thus if

$$W = X \pm Y \pm Z \pm \text{ etc.}, \qquad \ldots\ldots(4)$$

* This is the ratio of the sampling variances of (*a*) the mean deviation, and (*b*) root mean square deviation estimates of c, in large samples. I do not know from what source Gosset obtained these figures. The full value of the standard error of the mean deviation for samples of any size from a normal population was first derived, I believe, by Helmert (1876), but Gosset could not have known of this paper.

and E, e, f, g, etc., are the probable errors (or alternatively the moduli or the mean errors) of W, X, Y, Z, ... respectively, then Airy gives the law

$$E^2 = e^2 + f^2 + g^2 + \dots \qquad \dots\dots(5)$$

Gosset had noticed in certain cases he had met with that the result $E^2 = e^2 + f^2$ did not hold, as it should according to this law, for both $W = X + Y$ and $W = X - Y$. In other words he found that if W, X and Y are measured from their means there was very considerable difference between *Sum* $(X+Y)^2$ and *Sum* $(X-Y)^2$. He concluded that when this was the case it was a sign of the existence of a correlation between the variables. Thus he was feeling his way towards the fundamental relations (1) and (2) of p. 212 above, but he had not yet been introduced to the correlation coefficient.

The concluding remarks of the *Report* are interesting:

> We may point out that, although the proof of the law (of Error) rests on higher mathematics, the application of it only demands quite simple algebra. We have been met with the difficulty that none of our books mention the odds, which are conveniently accepted as being sufficient to establish any conclusion, and it might be of assistance to us to consult some mathematical physicist on the matter.

This last difficulty was repeated in the summary which contains the sentence:

> Explains that we have no information of the degree of probability to be accepted as proving various propositions, and suggests referring this question to a mathematician.

It is curious perhaps that Gosset should have felt at first that a mathematician was needed to solve this particular problem, which is just the point which the mathematician would now consider that the practical man must answer.* As we shall see in a moment he changed his view, but it seems to have been uncertainty on this question which led almost at once to that important contact between Gosset and Karl Pearson. A minute of March 1905 added to the printed *Report* indicates that arrangements for this meeting are to be made.

The interview was arranged through Vernon Harcourt, a chemistry don at Oxford whose pupil Gosset may have been and who perhaps got into touch with Pearson through Weldon, who was then Professor of Comparative Anatomy at Oxford. The opportunity for a meeting came about 12 July 1905 when Pearson was spending his long vacation at East Ilsley in Berkshire and Gosset bicycled over from his father's house at Watlington, preceded by a list of questions from which the following paragraphs are taken:

> (1) *My original question and its modified form.* When I first reported on the subject, I thought that perhaps there might be some degree of probability which is conventionally treated as sufficient in such work as ours and I advised that some outside authority should be consulted as to what certainty is required to aim at in large scale work. However it would appear that in such work as ours the degree of certainty to be aimed at must depend

* I have, however, heard of another very recent case where an industrialist considered that it was the mathematical statistician's job to suggest the appropriate odds to use.

on the pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment. This is one of the points on which I should like advice.

(2) *Another problem.* I find out the P.E. of a certain laboratory analysis from n analyses of the same sample. This gives me a value of the P.E. which itself has a P.E. of P.E.$/\sqrt{2n}$. I now have another sample analysed and wish to assign limits within which it is a given probability that the truth must lie. E g. if n were infinite, I could say "it is 10 : 1 that the truth lies within 2·6 of the result of the analysis". As however n is finite and in some cases not very large, it is clear that I must enlarge my limits, but I do not know by how much.

(3) *What is the right way to establish a relationship between sets of observations?* I use the following method when endeavouring to establish a relationship between sets of observations, but I have reason to suppose that it is not a good way and would like criticism on my method and advice as to the proper way. Suppose observations A and B taken daily of two phenomena which are supposed to be connected. Let A_1, A_2, A_3, etc. be the daily A observations and let B_1, B_2, B_3, etc. be the daily B observations. (I reduce the B observations if necessary or increase them by multiplying by a constant so that the P.E. of the A and B is about the same.) Then I form two series A_1+B_1, A_2+B_2, etc. and A_1-B_1, A_2-B_2, etc. and find the P.E. of each of the new series. If they are markedly different, it is clear (sufficient observations being taken) that the original series A and B are connected and proceed to attempt to find it quantitively. I cannot however at present find the P.E. of my results, nor can I be quite sure how great a difference between the P.E.'s of the sum and difference series is necessary to shew the connection.

(4) *What books would be useful?* When you talk with me you will doubtless find out many other points on which I require enlightenment and could perhaps recommend me some books on the subject.

The solution of "*another problem*" was to be given $2\frac{1}{2}$ years later in Gosset's paper on "The probable error of a mean" (2). The method described in paragraph (3) is interesting. I do not know exactly how Gosset attempted to measure the relationship quantitively, but if, as would seem natural, he compared the difference between $\Sigma(A+B)^2$ and $\Sigma(A-B)^2$ with their average, then by adjusting the scale so as to make the P.E.'s of A and B approximately the same, he had secured a maximum value for this ratio, and therefore presumably minimized the risk of overlooking a relationship. For

$$\frac{\Sigma(A+B)^2-\Sigma(A-B)^2}{\frac{1}{2}\{\Sigma(A+B)^2+\Sigma(A-B)^2\}}=\frac{4r_{AB}\sigma_A\sigma_B}{\sigma_A^2+\sigma_B^2},$$

which, for a given value of r_{AB}, has a maximum value of $2r_{AB}$ when $\sigma_A=\sigma_B$. One feels that, given a little more time, with his unerring instinct for reaching the best solution, Gosset would have found for himself Galton's correlation coefficient, just as he was later to rediscover Poisson's limit to the binomial and Helmert's distribution of a squared standard deviation.

Among Pearson's rough jottings written down for Gosset at the interview is the basic formula that he needed,

$$\sigma_{A\pm B}^2=\sigma_A^2+\sigma_B^2\pm 2r\sigma_A\sigma_B$$

(with the letter r doubly underlined), the probable error formula for r and also references to a number of papers on the theory of statistics.

Gosset was a quick learner; the immediate results of this visit include a *Supplement* to the brewery *Report* of 1904, from which I have quoted, and a second *Report* on correlation dated 30 August 1905. In both of these the influence of new ideas received from Pearson is evident. The *Supplement* contains a warning that distributions may not always be normal, although in small sample problems "it is practically convenient to use a curve...which has been thoroughly investigated, of which the values have been tabled, and which in the majority of cases describes them 'within the error of random sampling'". His colleagues are also advised to use the standard deviation and not the mean error. The *Report* is headed "The Pearson Co-efficient of Correlation", and describes, with a numerical example, the method of calculating this coefficient, r, as well as the use of the regression straight line for prediction.

This idea of correlation, which in origin is of course Galton's rather than Pearson's, has more than once during the past fifty years brought with it a stimulus leading to fresh discovery. The conception, presented with all its novelty to minds which had hitherto only considered the perfect relationship of the physicist as a relationship which could be scientifically handled, has seemed to provide a key to the solution of a host of problems. The inspiration which Galton's discussion of correlation in his *Natural Inheritance* gave to Weldon and Pearson in the early nineties has often been referred to and, now, the introduction of the new ideas opened out fresh avenues of research to both Gosset and his colleagues. The crude method which Gosset had invented of examining the difference between $\Sigma(A+B)^2$ and $\Sigma(A-B)^2$ could be abandoned. It became possible to assess with precision the relative importance of the many factors influencing quality at different stages in the complicated process of brewing, and before long the methods of partial and multiple correlation were mastered and applied.* The *Reports* circulated within the brewery constantly quote correlation coefficients and their probable errors, while Gosset's rough notebooks of this date contain numerous correlation tables. Apart from the actual calculation of r, the idea of arranging data in a two-way table was possibly novel and certainly illuminating to the brewers.

It seems, however, to have been at once obvious to Gosset that the methods developed by Pearson and his co-workers for handling the large samples met with in biometric inquiries would probably need modification when applied to the problems of the brewery. In his *Report* on correlation of August 1905 he notes that "correlation coefficients are usually calculated from large numbers of cases, in fact I have only found one paper in *Biometrika* of which the cases are as few in number as those at which I have been working lately". He was dealing at this time with all the possible correlations between a number of characters for which 31 observations were available; in another problem only 10 observations

* A *Report* of Gosset's of June 1907 applies multiple correlation to prediction. The mathematical Appendix dated 27 September 1906 is stated to have been read through by Karl Pearson.

could be used. He gives a reason which, though faulty, is extremely interesting, for doubting the validity of the probable error formula for r in small samples. Thus, if r is an observed correlation from a sample of n individuals, he takes the ratio

$$\frac{\text{Deviation of } r \text{ from zero}}{\text{Probable error of } r} = \frac{r}{0\cdot6745\,(1-r^2)/\sqrt{n}} \qquad \ldots\ldots(6)$$

as a measure of the significance of the correlation, remarking that if the ratio is greater than $2\frac{1}{2}$ the odds are about 20 : 1 on the existence of a real relationship. He then says that if n be very small "I expect a larger ratio is required", and illustrates this by supposing that $r=0\cdot9$, $n=4$, when the probable error calculated as in (6) becomes 0·064 and the ratio is 14. "Yet", he remarks, "no one would claim any certainty from four experiments."

If we are asking whether an observed r is consistent with sampling from a population in which the correlation, say ρ, is zero, then the appropriate probable error is approximately $0\cdot6745/\sqrt{n}$ and not the value used in (6). Thus in Gosset's example the ratio is really 2·7 and not 14; as he was afterwards to show, it was not the standard error that was seriously at fault in testing significance when dealing with small samples, but the assumption of normality. For $n=4$, $\rho=0$, the distribution of r is rectangular. The faulty reasoning involved in the interpretation of equation (6) has been used again and again in statistical literature; the reason that in 1905 the difficulty had not caught the attention of the workers at the Biometric Laboratory was that they were dealing with large samples and, for these, the error involved is of relatively small consequence. It was Gosset, "naughtily" playing about with absurdly small numbers,* who stumbled on the inconsistency, although not at first understanding its reason. Here perhaps we may see the first illustration of the tremendous gain in clear thinking that has followed in statistics from an approach to the subject from the small-sample end. Also this is one of the many occasions on which Gosset was first on the spot.

There were other difficulties in application that he was already turning over in his mind. For instance, he wished to obtain a combined measure of the correlation between two characters measured on several varieties of the barley used for malting and he considered the possibility of taking deviations from variety means. "I hope to find out the limitations of this device at some later date", he reported. "I am using it and similar devices pretty freely...."

A point which may be of interest to industrial statisticians to-day is that the practical brewer of thirty years ago, as the practical engineer to-day, was objecting to the introduction into his reports of the statistician's term *population*, yet was unable to suggest an appropriate substitute. A footnote to the word population ran as follows: "This appears to be a general statistical term to

* Writing to Gosset on 17 September 1912 on the subject of the standard deviation, not correlation, Pearson remarked that it made little difference whether the sum of squares was divided by n or $(n-1)$, "because only naughty brewers take n so small that the difference is not of the order of the probable error!"

express a number of things or people of the same kind. We have tried to find a word in common use to express this, but have failed."

The *Report* closes with a characteristic piece of sound advice.

It must be borne in mind, however, that the better the instrument the greater the danger of using it unintelligently: it is more important than ever to think carefully in what way any connection may have arisen accidentally, and, more especially, any semi-constant variation must be treated with particular care.

Statistical examination in each case may help much, but no statistical methods will ever replace thought as a way of avoiding pitfalls, though they may help us to bridge them.

THE YEAR IN LONDON, 1906–7, AND THE WORK ON SMALL SAMPLES

Following a general practice of the brewery, Gosset was sent away from Dublin for a year's specialized study. He spent the greater part of this time either working at or in close contact with the Biometric Laboratory, where he arrived at the end of September 1906. During the year which had elapsed since he first met Karl Pearson he must have given a great deal of time and thought to the application of current statistical methods to the type of experimental and routine data analysed in the brewery. He was now anxious to obtain Pearson's opinion on the work he had already done and to ask his advice on a number of unsolved questions. Probably he had already realized that the most important problem on which he required further information was the behaviour of frequency constants in small samples. In a letter written to a friend at the brewery on 30 September 1906, just after his arrival, he outlines, however, only a modest programme:

Then he [K. P.] proposes to give me a room to work in, that I should attend his lectures, and become as far as possible accustomed to the calculations, etc., of his department. I had a long talk with him, and told him the lines I had been going on in the Hops..., and he seemed to consider that I had been over most of the ground, but points soon cropped up which showed him the necessity for going deeper. I think that from what he said I am more or less on the right lines so far; perhaps when the reports have been considered you might let me have a copy of each of them, to ask about anything which may have occurred to me by then about them. I think he would be very willing to give us advice on any points which crop up.

The first problem which he took up was of considerable practical importance in one department of the brewery activities: the question of the sampling error involved in counting yeast cells with a haemacytometer. In his paper (1) published early in 1907 he derived afresh Poisson's limit to the binomial distribution, namely,

$$e^{-m}\left\{1+m+\frac{m^2}{2!}+\dots+\frac{m^r}{r!}+\dots\right\}, \qquad \dots\dots(7)$$

and showed by a comparison of the series with four sets of experimental results that it did represent well the observed distribution of cell counts in an investigation carried out under carefully controlled conditions. The paper should be read in conjunction with another that he wrote on the same subject twelve years later (9).

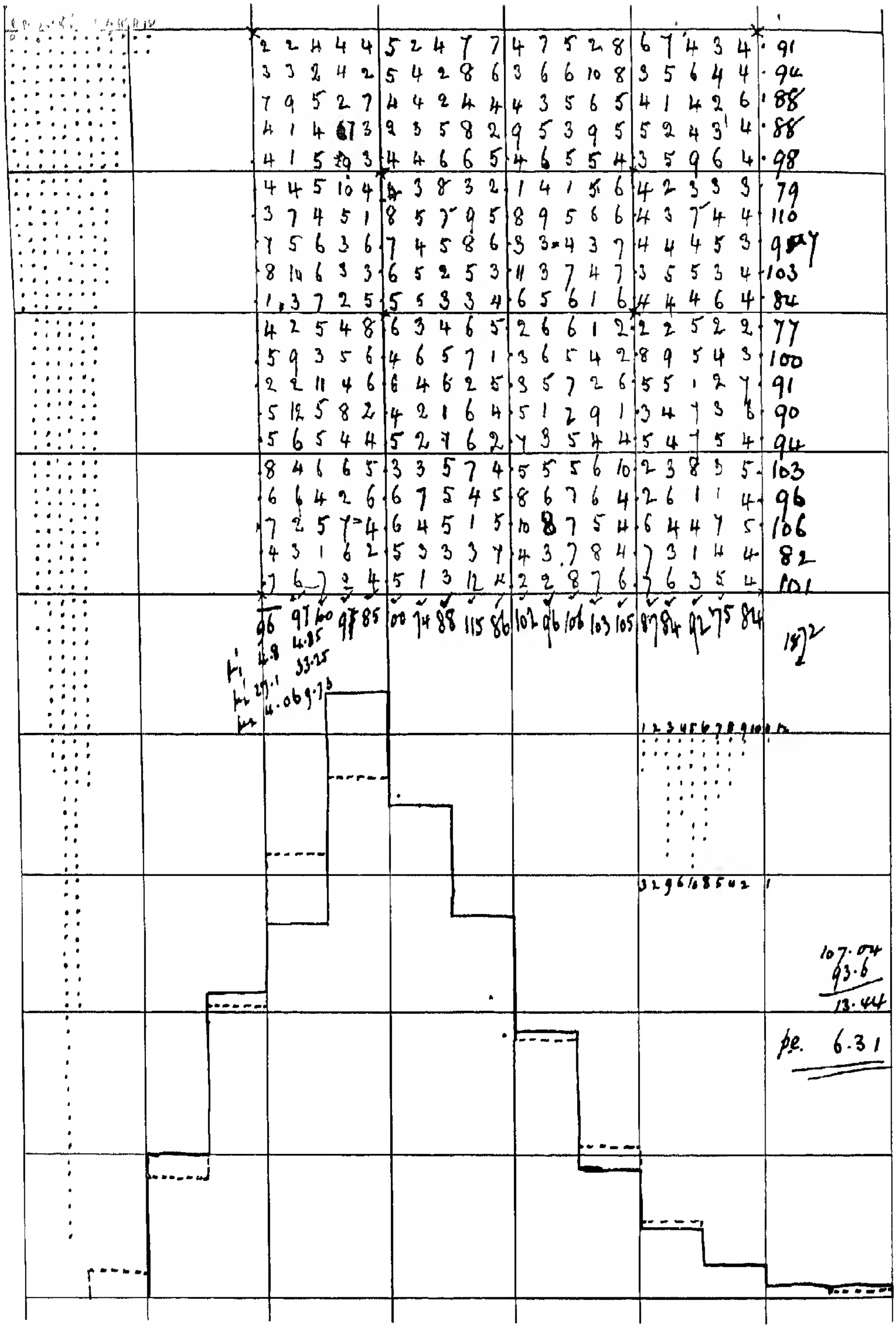
The derivation of the limiting form of the binomial was not in itself an achievement of any special difficulty; the series has been obtained independently from time to time by a number of investigators. But it was characteristic of "Student's" flair or, as he himself would have said, luck that when he had a practical problem to solve he should go straight to the correct solution; and that because it was a fundamental type of biological problem his research should have been of much greater value in the field of applied statistics than von Bortkiewicz's work, illustrated by fitting the Poisson series to suicides of German women and deaths of Prussian soldiers from the kicks of a horse.

I have reproduced in facsimile in Plate II two pages from Gosset's notebook containing the rough working for this paper. The experimental data are those of the series IV (see his p. 357). They are quoted also as an example of a Poisson distribution by R. A. Fisher in *Statistical Methods for Research Workers* (1938, p. 58). The left-hand page contains the 400 individual yeast cell counts and the resulting frequency distribution and histogram; the right-hand page shows the calculation of the mean, m, as well as the theoretical series and the derivation of χ^2. The expression $N/\surd(2\pi mq)$ (or $N/\surd(2\pi nq)$), where q is put equal to unity, is an approximation to the frequency in the group containing the mean. In the notes, Gosset seems to have reached this result by a rather lengthy method, but it can be obtained by putting $r=m$ in the general term of series (7) and using the first order term in Stirling's approximation to $m!$ No reference to this comparison was made in the published paper. A few figures, which are in pencil in the note-book, appear to be in Pearson's hand; e.g. the theoretical frequencies 3·712, 17·37 and 40·65 as well as the three terms of the Poisson series at the bottom of the page. They were jottings made no doubt by K. P. on one of his daily "rounds" of the laboratory.

A good part of the work on Gosset's second paper on "The probable error of a mean" (**2**) was also carried out during his year in England; with it is closely associated his third paper on the "Probable error of a correlation coefficient" (**3**), as both were supported by the same piece of experimental sampling. I have already referred to Gosset's doubts regarding the distribution of r in small samples; since in the brewery work a mean value had often to be estimated from eight or ten determinations he also felt uneasy about the applicability to such work of accepted theory regarding the distribution of the mean and the standard deviation. A letter written on 12 May 1907 to a colleague in Dublin shows him to be in the middle of his investigation. After dealing with some points about the significance of differences* he adds:

> Herewith my answer to your questions. I hope it is quite clear, but I am afraid I rather increase the difficulties when I try to explain anything as a rule.

* There is a reference to that long-standing difference of opinion regarding n and $n-1$, in the following sentence: "When you only have quite small numbers, I think the formula we used to use for the P.E. ($\surd\{\Sigma(x^2)/(n-1)\}\times 0{\cdot}6745$) is better, but if n be greater than 10 the difference is too small to be worth taking the extra trouble." Here K. P. and Airy were in disagreement.



2	2	4	4	4	5	2	4	7	7	4	7	5	2	8	6	7	4	3	4	91
3	3	2	4	2	5	4	2	8	6	3	6	6	10	8	3	5	6	4	4	94
7	9	5	2	7	4	4	2	4	4	4	3	5	6	5	4	1	4	2	6	88
4	1	4	7	3	2	3	5	8	2	9	5	3	9	5	5	2	4	3	4	88
4	1	5	9	3	4	4	6	6	5	4	6	5	5	4	3	5	9	6	4	98
4	4	5	10	4	6	3	8	3	2	1	4	1	5	6	4	2	3	3	3	79
3	7	4	5	1	8	5	7	9	5	8	9	5	6	6	4	3	7	4	4	110
7	5	6	3	6	7	4	5	8	6	3	3	4	3	7	4	4	4	5	3	97
8	10	6	3	3	6	5	2	5	3	11	3	7	4	7	3	5	5	3	4	103
1	3	7	2	5	5	5	3	3	4	6	5	6	1	6	4	4	4	6	4	84
4	2	5	4	8	6	3	4	6	5	2	6	6	1	2	2	2	5	2	2	77
5	9	3	5	6	4	6	5	7	1	3	6	5	4	2	8	9	5	4	3	100
2	2	11	4	6	6	4	6	2	5	3	5	7	2	6	5	5	1	2	7	91
5	12	5	8	2	4	2	1	6	4	5	1	2	9	1	3	4	7	3	6	90
5	6	5	4	4	5	2	7	6	2	7	3	5	4	4	5	4	7	5	4	94
8	4	6	6	5	3	3	5	7	4	5	5	5	6	10	2	3	8	3	5	103
6	6	4	2	6	6	7	5	4	5	8	6	7	6	4	2	6	1	1	4	96
7	2	5	7	4	6	4	5	1	5	10	8	7	5	4	6	4	4	7	5	106
4	3	1	6	2	5	3	3	3	7	4	3	7	8	4	7	3	1	4	4	82
7	6	7	2	4	5	1	3	12	4	2	2	8	7	6	7	6	3	5	4	101
96	97	100	97	85	100	74	88	115	86	102	96	106	103	105	87	84	92	75	84	1872

Page from Gosset's notebook containing the analysis of haemacytometer counts. (Left-hand page.)

						χ^2	
			Calc ~~3.704~~	3.712	−4	4.00	vp q
0	0						
1	20	20	17	17.37	+3	.53	$(1-vp)$
2	43	86	41	40.65	+2	.10	$(P+Q)=1$
3	53	159	63	63	−10	1.59	
4	86	344	74	74	+12	1.95	
5	70	350	70	69	0		
6	54	324	54	54	0	$1+\frac{1}{1}+\frac{1}{1.2}+\frac{1}{1.2.3}+\ldots$	
7	37	259	36	36	+1	.03	
8	18	144	21	21	−3	43	
9	10	90	11		+1	.09	
10	5	50	5		0	$5^2 = 25$	
11	2	22	2		0		
12	2	24	1		+1	1.00	
	400	1872				9.72	

about .64

−.2032498
2.602060
.569562

Average number 4.68   from $(.9525+.0475)^{98.53}$ $\chi^2 = 9.28$

Calculating the number of mean cases on formula $\frac{N}{\sqrt{2\pi n q}}$ & putting $q = 1$ for this purpose we get the number to be 73.8 (interpolating) a fair agreement between 86 @ 4 & 70 @ 5 & 4.68 & not 75 or so

(4.68) $1 + nq$

$400\left(\frac{399}{400} + \frac{1}{400}\right)^{1872}$ $\left(\frac{1872}{400}\right)\times 400$ $1 + \frac{1872}{400}$

$1 + 4.68$

$Me^{-m}\left\{1 + m + \frac{m^2}{12}\right.$

Page from Gosset's notebook containing the analysis of haemacytometer counts. (Right-hand page)

What I have written on the back is true for large samples, and approximately so for small, and is the accepted theory. My work on small numbers may or may not modify it. We shall know later....

I go up to K.P.'s lectures from here [The Ousels, Tunbridge Wells] and on other days work at small numbers: a greater toil than I had expected, but I think absolutely necessary if the Brewery is to get all the possible benefit from statistical processes.

There could be no better illustration than these last sentences of the way in which Gosset's best work was called forth in the service of his firm.

The contents of the paper on the probable error of the mean are too well known to require more than a brief summary. Starting with a sample of n observations, $x_1, x_2, \ldots, x_n$, from a normal population with standard deviation σ and mean at the origin for x, Gosset obtained the sampling moments of $s^2 = \Sigma(x-\bar{x})^2/n$, where $\bar{x}$ is the sample mean. He showed that these moments corresponded exactly with those of a Pearson type III curve and hence inferred that the curve representing the sampling distribution of s^2 must almost certainly be

$$y = \text{constant} \times \sigma^{-n+1}(s^2)^{\frac{1}{2}(n-3)}\, e^{-ns^2/2\sigma^2}. \qquad \ldots\ldots(8)^*$$

He then showed that the correlation coefficient between $\bar{x}^2$ and s^2 was zero and, making the assumption (which does not necessarily follow though in fact it is true in this case) that this meant that $\bar{x}$ and s were absolutely independent, he deduced the probability distribution of $z = \bar{x}/s$ as

$$p(z) = \text{constant} \times (1+z^2)^{-\frac{1}{2}n}. \qquad \ldots\ldots(9)$$

He considered the properties of this curve,† gave a table of its probability integral for $n=4$ to 10 and examined its approach to a normal curve with standard deviation $1/\sqrt{(n-3)}$. He next compared the distributions (8) and (9) with the results of a sampling experiment for the case $n=4$ and finally illustrated the use of his results on four examples.

When two years ago the question of the photographic reissue of the paper had been suggested to meet a continued demand for offprints, Gosset wrote to me describing it as now "rather a museum piece". That is true, though perhaps in a different sense than he meant. It is a paper to which I think all research students in statistics might well be directed, particularly before they attempt to put together their own first paper. The actual derivation of the distributions of s^2 and z, or of $t = \sqrt{(n-1)}\, z$ in to-day's terminology, has long since been made simpler and more precise; this analytical treatment need not be examined carefully, but there is something in the arrangement and execution of the paper which will always repay study.

In the first place, in the Introduction and Conclusions we find an excellent illustration of Gosset's wise advice given to a beginner in the art of composition. "First say what you are going to say, then say it and finally end by saying that

* That this result had previously been derived by Helmert (1876), English-speaking statisticians were quite unaware till many years later.

† There are some minor errors in §§ IV and V of the paper.

you have said it."* The main part of the paper, the "saying it", is divided clearly into headed sections. The adequacy of the assumptions on which the mathematical theory rests is tested by a piece of experimental sampling; this test being satisfactorily passed, computed tables required for application are given and finally a number of well-chosen examples illustrate the purpose of the inquiry.

Before considering some other notable features of the paper and attempting to assess its influence on later work, it is important to see just what was the main purpose of the inquiry that its author had in mind. As usual with him, this was simple and practical. Having n observations, he wished to know within what limits the mean of the sampled population—the "true result" of the 1904 *Report*—probably lay. His solution involved a tacit introduction of the method of inverse probability, but I do not think he ever tried to put this into precise terms.† Thus the last sentence on the first page of the paper runs as follows:

> The usual method of determining the probability that the mean of the population lies within a given distance of the mean of the sample, is to assume a normal distribution about the mean of the sample with a standard deviation equal to $s/\sqrt{n}$, where s is the standard deviation of the sample, and to use the tables of the probability integral.

The results of the present investigation meant to Gosset that he could now assume in small samples a z-distribution for the population mean about the sample mean, the scale now being the sample standard deviation, s. In his examples he uses the z tables, not to test the hypothesis that the population mean is zero or has some other specified value, but to find the odds that this mean lies within specified limits, e.g. between 0 and ∞, that is to say is positive. Take for instance his *Illustration* I (pp. 20–1); the average number of hours of sleep gained by ten patients treated with *D. hyoscyamine hydrobromide* is $\bar{x}=0{\cdot}75$ while the standard deviation is $s=1{\cdot}70$. If we regard the population mean, say ξ, to be distributed about the sample mean 0·75 in the z-form, with a standard deviation of s, it follows that the chance that $\xi>0$ is the proportionate area under the z-curve between the ordinate at

$$z=\frac{0-0{\cdot}75}{1{\cdot}70}=-0{\cdot}44$$

and ∞. This is the same as the chance that $z<+0{\cdot}44$, which interpolation in his tables in the column $n=10$ shows to be 0·887. He therefore argued that the odds are 0·887 to 0·113 that the population mean ξ is positive, i.e. that the soporific will

* The advice was not originally Gosset's. Writing in 1934 he says: "This is a rule which we owe to A. J. (I think at second hand)." He then quotes the rule and adds, "It does make things so much easier for everybody concerned, besides which 'what I tell you three times is true'"; the last words are those of the Bellman in *The Hunting of the Snark*.

† In his paper on the correlation coefficient written in the same year (**3**, p. 302) Gosset states definitely that a knowledge of the *a priori* probability distribution of the population correlation coefficient, R, is needed in order to determine "the probability that R...shall lie between any given limits".

on the average give an increase of sleep. While a somewhat loosely defined conception of inverse probability seems to underlie the argument, it will be seen that as far as the practical consequences go, Gosset had reached a result which we can hardly improve on 30 years later. It is true that, using the idea of the fiducial or confidence interval, some of us would word our statement of limits and probabilities a little differently so as to avoid any appeal to inverse probability, but as practical statisticians we must, I think, admit that our conclusions would be identical.

There are some other features of the paper which are interesting historically. Gosset remarks on p. 13 that before he succeeded in solving the problem analytically, he had endeavoured to do so empirically. The sampling experiment which he carried out for this purpose involved the drawing of 750 samples of 4 by means of shuffled slips of cardboard, from W. R. Macdonell's (1901) correlation table containing the distribution of height and middle-finger length of 3000 criminals. As far as I know this was the first instance in statistical research of the random sampling experiment which since has become a common and useful feature in a large number of investigations where precise analysis has failed. The results of this same experiment were used by Gosset in a number of later papers. On p. 16 he draws attention to a difficulty in the application of Pearson's χ^2-test of goodness of fit which was later to lead to R. A. Fisher's modification in terms of degrees of freedom. On p. 19 he gives reasons for believing that even when the population sampled is not normal the sampling distribution of z will be very little modified; this was a prediction which experimental and theoretical investigations carried out in recent years have confirmed.

Finally we may note the introduction of a difference in notation to distinguish between sample and population characters, viz. s for the sample and σ for the population standard deviation. The need for this distinction seems obvious to us to-day, but it is interesting to notice that it was only when attention was directed to the problem of small samples that statisticians grasped the clarification resulting from this innovation.

As the theory of mathematical statistics has developed, the significance of "Student's" test has been elaborated from many angles and deeper meanings associated with it than its author had ever dreamed of. This is a common feature of scientific progress, but as Neyman very appropriately remarked on a recent occasion (1937, p. 142): "The role of a rigorous scientific theory is frequently very modest and is reduced to explaining to the practical man—and this sometimes with a certain difficulty—how good is what he himself knew to be good long ago." To understand the reason for the historical importance that has rightly been associated with this paper, it is not however necessary to discuss the abstract conceptions of the mathematical statistician and their relation to forms of critical regions in hyperspace; it can be explained much more simply

than that. As Gosset wrote on the second page of the paper, referring to the inadequacy for certain purposes of the statistical technique available in 1908.

> There are other experiments, however, which cannot easily be repeated very often; in such cases it is sometimes necessary to judge of the certainty of the results from a very small sample, which itself affords the only indication of the variability. Some chemical, many biological, and most agricultural and large scale experiments belong to this class, which has hitherto been almost outside the range of statistical inquiry.

It is probably true to say that this investigation published in 1908 has done more than any other single paper to bring these subjects within the range of statistical inquiry; as it stands it has provided an essential tool for the practical worker, while on the theoretical side it has proved to contain the seed of new ideas which have since grown and multiplied an hundredfold.

The sampling experiment used to test the accuracy of the theoretical distributions of s^2 and z was also planned to throw light on the distribution of the correlation coefficient r, in very small samples. In this second problem (3) Gosset was forced to rely much more on his empirical approach than before, since the mathematical solution lay beyond his powers. In suggesting the probable form of the distribution of r when sampling from a population in which the two variables were uncorrelated (i.e. $R=0$)* he could get no clue from known values of moments as in the case of s^2. He started from the following basis: (a) the distributions must be symmetrical about $r=0$ and be limited within the range -1 to $+1$; (b) he had available the distributions of r found from his experiment for 745 samples of 4 and 750 samples of 8; (c) of these, he noticed that the former was approximately rectangular.

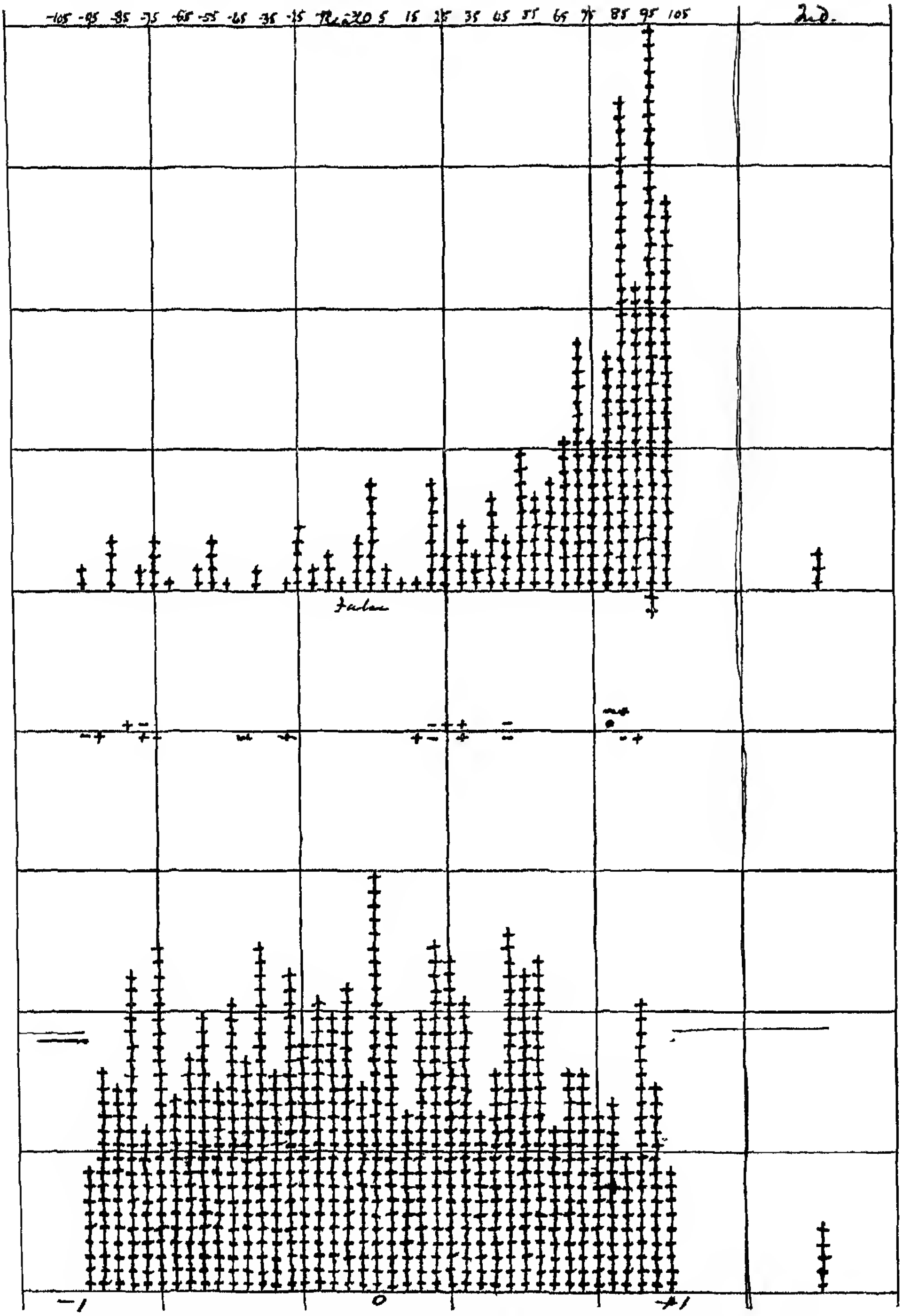
As in the case of s^2, his training at the Biometric Laboratory naturally suggested that he should try to use a Pearson curve for the unknown distribution; a type II curve was the only one suitable, and therefore in his own simply expressed phrase, "working from $y=y_0(1-x^2)^0$ for samples of 4 I guessed the formula

$$y=y_0(1-x^2)^{\frac{1}{2}(n-4)}\text{''}. \qquad \ldots\ldots(10)$$

He then showed that for $n=8$ this formula represented his empirical sampling distribution very well, and pointed out that the result agreed with large sample theory, since the standard deviation $\sigma_r=1/\sqrt{(n-1)}$ would equal Pearson and Filon's value of $(1-R^2)/\sqrt{n}$ when $R=0$ and $n\to\infty$. He also gave the correct limiting result, which he had been able to establish for any R, when $n=2$, suggesting that this might furnish a clue for the distribution when $n>2$. It was a brilliant piece of guessing and all the more striking because of the forceful way in which the supporting evidence was marshalled.

In the case where the population correlation, R, was not zero Gosset provided three empirical sampling distributions for the cases $R=0{\cdot}66$ and $n=4$, 8 and 30

* He used R for the population correlation; the notation, ρ, seems to have been first used by H. E. Soper (1913).



Distribution of the correlation coefficient in samples of 4, tabled in Gosset's notebook.
Above, $R = 0\cdot 66$; *below*, $R = 0$.

He also set out very clearly the conditions which his work showed must be satisfied by the true distribution. "I hope", he concluded, "they may serve as illustrations for the successful solver of the problem". Six years later R. A. Fisher was able to demonstrate the substantial accuracy of all Gosset's predictions both in the r and the z paper.

In the notebook containing the original samples of 4 from Macdonell's correlation distribution, there are given what I think must be the original distributions built up by Gosset as he tabled his calculated values of r. Two of these are shown in facsimile in Plate III ($n=4$, $R=0{\cdot}66$ and $n=4$, $R=0$). It is hard to believe that Gosset did not experience a very pleasurable excitement as these distributions gradually took shape on the paper, for he was exploring a region entirely unmapped and the discovery of the rectangular distribution in the case when $R=0$ must have been a complete surprise.*

One of the curious things that must strike us now about these two papers of Gosset's (2, 3) is the small influence that their publication had for a number of years on current statistical literature and practice. The z-test was used in the brewery at once, but I think very little elsewhere for probably a dozen years. Perhaps because he realized that it showed how little reliability could be placed on a correlation coefficient based on small numbers, Gosset does not seem to have recommended the use of the r-test even to his colleagues and he made no tables of the probability integral for the distribution (10). I have come across, however, one reference to the work in a letter of 3 April 1912 to E. S. Beaven, in which the following remarks occur:

> By the way, don't be *too* cock-a-hoop about your 0·95 correlation with 7 cases. Such a thing might occur more than once in a hundred trials of 7 cases, even if there were no correlation. (I haven't got tables to evaluate
>
> $$\int_{\sin^{-1}0{\cdot}95}^{\frac{1}{2}\pi}\cos^4\theta\,d\theta \Big/ \int_0^{\frac{1}{2}\pi}\cos^4\theta\,d\theta,$$
>
> but you get that fraction of N at each end 0 95 or over in N trials); and I guess its about 2 % at each end.† All the same it seems very reasonable to suppose that it is right.

From Gosset's point of view, he had developed the tools which he needed for practical application in Dublin and he was not primarily interested in their wider use. If Pearson failed to realize the importance of the work and did not assimilate the results into current practice and teaching, it was because he too was mainly interested in what appeared to be of value in the research investigations of his laboratories. To him all small sample work was dangerous and should be avoided. But it would be wrong to suppose that there was a lack of sympathy between the two; except at a far later stage when opposite views over z found their way into print, Pearson's attitude towards Gosset's small sample

* Yet Mrs Gosset, who was helping him at the time, writes: "Whatever thrill he may have got out of that experiment he showed nothing whatever of it, and his amanuensis never realized that there was anything original about it!"

† Gosset was wrong here. The fraction is actually 0·001.

work was one of humorous protest, well conveyed in the quotation I have given about "naughty brewers" who take n too small (p. 218 above). The readiness with which he would talk to Gosset over his problems and at times refer to him on matters of difficulty shows how highly he rated his ability and insight Although Gosset launched off along independent lines of investigation directly he had mastered the elements of statistical theory, it is clear that he owed a great deal to the early guidance that he received in London. In the first place he had that very great advantage of being freed for a year from his official duties and of spending that time in close contact with persons who were enthusiasts in the study of statistics. Although, as he wrote at a later date, "I am bound to say that I did not learn very much from his [K. P.'s] lectures; I never did from anyone's and my mathematics were inadequate for the task", he obtained from the Biometric Laboratory a number of things which were not to be found in Airy or Merriman: the theory of correlation, the χ^2-test, and above all Pearson's system of frequency curves. It is doubtful for instance if he could have reached the distribution of s^2, and hence that of z, if he had not had available for use Pearson's type III curve.

After his year in London was over Gosset kept in close touch with Pearson for 29 years, and to his intimate friends would speak with admiration of his teacher. Some sentences which he spoke at the opening meeting of the Industrial and Agricultural Research Section of the Royal Statistical Society in November 1933 were composed, I know, with this aspect of the relationship between professor and student in mind:

> Another point arises from the peculiar nature of statistics. It is impossible to apply statistical methods to industry or anything else unless one has a certain amount of intelligent experience as a background. That works both ways. The practical man has to go and talk to his Professor partly in order that the Professor himself should share his experience. ...The whole art of statistical inference lies in the reconciliation of random mathematics with biassed samples. Every new problem has some fresh kind of bias and might contain some new pitfall. The only way not to fall into these pitfalls is to talk over the problem with some intelligent critic; and so the practical man, if he is not entirely foolish, talks over his problems with the Professor, and the Professor does not consider himself to be a competent critic unless he has had some experience of applying the statistics to industry and has learned the difficulties of that application.

Miscellaneous papers, 1909–21

Before considering the very important part that Gosset played in the development of agricultural experimentation, it is desirable to give a brief account of six papers on a variety of subjects which were published in *Biometrika* between 1909 and 1921.

(i) The first of these papers on "The distribution of the means of samples which are not drawn at random" (4, 1909) dealt with one aspect of that theme which,

as I have already mentioned, runs through so much of his work. He had realized at an early date how frequently there existed a correlation between successive observations either in time or space. Thus if x and y are two contiguous observations it would follow that

$$\sigma^2_{x+y} > \sigma^2_x + \sigma^2_y > \sigma^2_{x-y}.$$

Hence if x and y were successive duplicate chemical analyses of the same quantity their mean would be less reliable than we should expect on the usual theory of random sampling. On the other hand were x and y the yields from plots of two different cereals which were to be compared, by placing the plots side by side in space, the difference $x-y$ would be more reliable than on the classical error theory. In this paper he considers the distribution of the mean not of two but of n observations, so selected that they are correlated, i.e. more like one another than individuals randomly selected from the population. It is the problem of fraternities which Pearson had termed homotyposis in his biometric work. Gosset gave the second, third and fourth moments of the sample mean, the second having the value

$$M_2 = \frac{\sigma^2}{n}\{1+(n-1)\rho\}, \qquad \text{......(11)}$$

where σ is the population standard deviation of x and ρ the correlation between the x's in a sample, which Fisher has termed the intraclass correlation. From the values of the third and fourth moments he deduced that in general it was likely that the distribution of the mean would tend to normality less rapidly than when $\rho=0$.

From the practical point of view he was concerned to warn the chemist that "repetition of analyses in a technical laboratory should never follow one another, but an interval of at least a day should occur between them. Otherwise a spurious accuracy will be obtained which greatly reduces the value of the analyses".

(ii) The next paper (6) published in 1913 dealt with "The correction to be made to the correlation ratio for grouping", an investigation no doubt connected with Pearson's work (1913) on the same subject published in the same number of *Biometrika*.

(iii) Volume x of *Biometrika* (1914) contains a short note on "The elimination of spurious correlation due to position in time or space" (7). In this, Gosset showed that the difference correlation method used by F. E. Cave (1904) and R. H. Hooker (1905) could be extended to differences of higher order than the first. This paper was the basis of later investigations on the variate difference correlation method.

(iv) In 1917 (8) Gosset published an extension of his tables of the probability integral of z; the range covered now ran from $n=2$ to $n=30$. In the intro-

ductory remarks he again gave advice "as to the best way of judging the accuracy of physical or chemical determinations". He wrote:

> After considerable experience, I have not encountered any determination which is not influenced by the date on which it is made; from this it follows that a number of determinations of the same thing made on the same day are likely to lie more closely together than if the repetitions had been made on different days. It also follows that if the probable error is calculated from a number of observations made close together in point of time much of the secular error will be left out and for general use the probable error will be too small. Where then the materials are sufficiently stable, it is well to run a number of determinations on the same material through any series of routine determinations which have to be made, spreading them over the whole period.

(v) Gosset's paper of 1919 (9) on "An explanation of deviations from Poisson's law in practice" answered some questions regarding the relation of this series to the positive and the negative binomial raised by Lucy Whittaker (1914) in a paper published five years earlier from the Biometric Laboratory. Since the rather severe criticisms of the latter paper directed against the applications of the Poisson law made by Bortkiewicz and Mortara might have discouraged its use in other directions, Gosset pointed out that the object of his own earlier paper (1) was to give the user of the haemacytometer a guide to the error of his count. From this first practical point of view it made little difference whether, theoretically, the better fitting distribution was a positive or negative binomial, although as a further point it was of interest to consider what such departures implied if the data were sufficient to establish them.

(vi) The final paper (10) of this group on "An experimental determination of the probable error of Dr Spearman's correlation coefficients", was written in the first instance for reading at one of the early meetings (13 December 1920) of the newly formed Society of Biometricians and Mathematical Statisticians. Gosset had many years before realized the value of the method of rank correlation in assessing quickly the order of relationship between two short series of numbers. Probably while working at the Biometric Laboratory he had developed the proof quoted by Pearson (1907, p. 13), that the standard error of the coefficient

$$\rho = 1 - \frac{6\Sigma D^2}{n(n^2-1)} \qquad \ldots\ldots(12)$$

is $1/\sqrt{(n-1)}$, in the case of independence in the population. In a Report written in 1911 for his colleagues in the brewery he illustrated the use of the method and gives what is substantially the correction for "ties" described in the present paper of 1921. Apart from the publication of this correction, the paper is of interest because Gosset again made use of his sampling experiment of 1907. For the 375 samples of 8 from a population having correlation 0·66 he calculated both of Spearman's rank correlation coefficients, in their raw and corrected form and, in the case of his 100 samples of 30 added Sheppard's estimate of correlation obtained from a median fourfold division. He uses these results to make a number of comparisons between the methods, in particular paying

regard to the amount of additional sampling needed if one of these more rapid methods of "assay" is to give as reliable an estimate of the population correlation coefficient as that obtained from the usual product-moment formula. He concludes by suggesting to mathematicians a problem which has still remained unsolved, that of determining the sampling distribution of the rank coefficient of equation (12) above, in random samples from a bivariate normal population, in which the correlation is not zero.

THE APPLICATION OF STATISTICAL METHOD TO AGRICULTURAL PLOT EXPERIMENTS

It is a feature commonly noticeable in the advance along any new line of scientific inquiry that the first steps in that progress are made hesitatingly and with difficulty, accompanied by much trial and error; and then after many years of what seems, looking back, to have been a painfully slow advance to an obvious goal, a stage is reached where the way forward has been almost cleared so that the introduction, perhaps, of some new tool or some fresh personality leads to a rapid advance into fresh country. In later years the casual student may well attribute the beginning of an epoch to that moment of rapid advance, partly because few records of the earlier struggle have found their way into print and partly because the later workers themselves have hardly realized the amount of thought that has gone into the creation of ideas which have formed the groundwork of their own further progress.

The history of the introduction of statistical methods in the planning and interpretation of agricultural experiments provides an illustration of these points. The large extension of technique with the accompanying stimulus to scientific planning which followed R. A. Fisher's introduction of the methods of analysis of variance in the years following 1923, may have caused the present-day statistician to overlook the essential pioneer work of the preceding years, without which it is certain that the later advance would have been impossible.* It therefore seems appropriate to take this opportunity of giving rather special attention to this aspect of Gosset's contribution to statistics and to do so by following out the gradual stages by which he advanced from simple beginnings to the analysis of a balanced block experiment.

A number of persons contributed to this early work and, as is often the case when methods of attack are in an imperfect or trial stage, ideas were worked out in correspondence or by word of mouth rather than in print. The brewery, as a very large consumer of barley, was naturally interested in agricultural problems and in particular in certain large-scale experiments undertaken in Ireland under the supervision of the Irish Department of Agriculture. Gosset was not, however, concerned with giving advice in these experiments till a

* Fisher himself has on many occasions paid a warm tribute to the help he received both from "Student's" published work and from correspondence and discussion.

number of years after he had specialized in statistics, and I think his first real interest in agricultural work arose from his contact with E. S. Beaven, who as a maltster was from time to time in Dublin on official business. Beaven had started experimental work in the nineties and about 1905 approached Gosset for an interpretation of apparently anomalous results, afterwards seen to be due to interference, that he found in comparing the yields of two varieties of barley in his "cage" at Warminster. From that date until Gosset's death there was a continuous flow of correspondence between them in which ideas were exchanged and thrashed out, and the more mathematical approach of the younger man was influenced by the practical experience of his older friend.

It will be noticed that three out of Gosset's four illustrations in the paper on the probable error of the mean (2) deal with agricultural topics; the data were taken from published accounts of Woburn farming experiments and Gosset shows how, by taking appropriate differences and using his z-test, a more precise interpretation of such results could be obtained than had hitherto seemed possible. Beaven was in touch with the agricultural work both at Rothamsted and Cambridge and it was no doubt owing to his report of Gosset's keen interest in these problems that both of those classical papers by Wood & Stratton (1910) and by Mercer & Hall (1911), dealing with the analysis of what we now term uniformity trial data, passed through Gosset's hands before publication. The first was only "an affair of a day or two's glancing at" after which he "made one or two suggestions, most of which were quite rightly turned down as being too refined for the purpose".* But in the second case he "brooded over the paper for months", and made suggestions which were incorporated, as well as adding an Appendix (5). If we compare the two statistical contributions, that of Stratton to the first paper and that of Gosset to the second, it is possible, I think, to see without difficulty the latter's special contribution to the subject. Stratton is following the approach of the classical theory of errors, which he had learnt and applied as an astronomer; he shows that variation in plot yields can be represented by the error curve and hence that the results of that theory regarding the probable error of a mean are applicable. These results are used to show the relation of size and number of plots (or animals) to the reliability of the results. No reference is made to "Student's" paper of 1908.

Gosset, writing his Appendix a year later, brings to the problem the added insight that he has gained from an understanding of correlation theory and from much discussion of the Warminster results with Beaven. He shows how it is possible to bring the changing fertility level or "patchiness" of the experimental field into service (*a*) by scattering the varieties to be compared in small plots over the field, and (*b*) then taking as the statistical variable for analysis the difference between the characteristics of two varieties on neighbouring plots. Thus the standard error, by way of formula (2), p. 212 above, can be very much

* These quotations come from a letter of 4 June 1922 from Gosset to Beaven.

reduced. The illustration which he gives deals only with the case of two varieties A and B, and at this date he had probably not thought out a technique for dealing with more comparisons.

There is another point of difference that may perhaps be noted; Wood and Stratton by raising the question, "What is the probable error of a single field experiment?" seemed to suggest that it might be possible to determine a single value, σ, which it would be appropriate to apply to future experiments of a given type. Gosset however emphasized a rather different idea. He writes (5, p. 130):

> But, it will be asked, why take all this trouble? The error of comparing plots of any given size has been found by the authors of the paper, and all that has to be done is to apply this knowledge to the particular set of experiments.
>
> The answer to this is that there is no such thing as the absolute error of a given size of plot. We may find out the order of it, be sure perhaps that it is not likely to be less than (say) 5 per cent. nor more than 15 per cent.... but the error of a given size of plot must vary with all the external conditions as well as with the particular crops upon which the experiment is being conducted, *and it is far better to determine the error from the figures of the experiment itself; only so can proper confidence be placed in the result of the experiment.**

His own z-distribution was available, if the number of observations was scanty.

If the field were divided into m pairs of plots and x_i and y_i were the yield, say, of varieties A and B on contiguous ith plots, then Gosset's test for a difference in yield may be summarized as follows.

Write $d_i = x_i - y_i$ and $\bar{d} = \sum_i d_i/m$.

Calculate the ratio
$$z = \frac{\bar{d}}{\sqrt{\left\{\sum_i (d_i - \bar{d})^2/m\right\}}}, \qquad \text{......(13)}$$

and if $m \leqslant 10$ refer this to the z-tables (2, p. 19). Otherwise, if $m > 10$, since z has a standard deviation of $1/\sqrt{(m-3)}$, refer $z\sqrt{(m-3)}$ to Sheppard's tables of the normal probability integral.

In the years 1912 and 1913 at Beaven's suggestion plot experiments of similar design, each comparing eight varieties of barley, were carried out at three centres, viz. Warminster, Cambridge and Ballinacurra in Co. Cork. The experiments were carried out in cages, and there were twenty replications of each of the eight varieties in square-yard plots. The arrangement of the varieties in a "chess-board" pattern was effectively what we should now term balanced; a plan of one of the schemes has been shown in Gosset's paper of 1923 "On testing varieties of cereals" (11, p. 277) and I have reproduced a portion of this below, only adding some thicker rules to separate the different sets of eight plots.

Beaven suggested that the results might be analysed by using as a statistical

* The italics are mine.

variable the difference between (1) the yield on a plot of A, say, and (2) the mean yield for the eight varieties (including A) on the 9-plot area in which this A-plot lay at the centre.* This was a rough and ready procedure but, as Gosset pointed out, owing to correlation there would be difficulty in the statistical interpretation. The method which he preferred was a very natural extension of his difference method advocated in the case where there were only two varieties. He could still clearly use that method to compare any two of the eight varieties,

E 230·1	*B* 249·3	*G* 312·2	*D*	*A*	*F*	
D 255·9	*A* 222·6	*F* 218·7	*C*	*H*	*E*	
C 265·6	*H* 205·0	*E* 246·7	*B*	*G*	*D*	
B 265·9	*G* 236·7	*D* 295·8	*A*	*F*	*C*	*H*
A 236·5	*F* 210·4	*C* 291·1	*H* 223·9	*E*	*B*	*G*

Fig. 1.

say A and D, taking the corresponding pair of plots from each set of eight, and differencing the character measured, although the plots would not now be generally contiguous. This would mean that changes in soil fertility, etc. would make the comparison less accurate than before,† but that could not be helped if eight varieties were to be compared in a single experiment in place of two. He saw, however, that it was possible to compensate to some extent in another direction for this loss in accuracy, by getting a single combined estimate of error from all the $\frac{1}{2}n(n-1) = 28$ possible sets of differences between $n = 8$ varieties, a method which he described as "hotchpotching" the comparisons. The reasoning which he used in reaching his result may be set out as follows:

Let there be n varieties each repeated m times and denote by $d_{uv\cdot i}$ the difference obtained from the ith comparison of the uth and vth varieties ($i = 1, 2, \ldots, m$) and by $\bar{d}_{uv}$ the mean of these m differences. Thus in Fig. 1, if u and v stand for varieties A and D, respectively, then

$$d_{uv\cdot 1} = 236{\cdot}5 - 255{\cdot}9 = -19{\cdot}4, \quad d_{uv\cdot 2} = 222{\cdot}6 - 295{\cdot}8 = -73{\cdot}2, \text{ etc.}$$

To obtain a common estimate of the standard deviation of differences, say σ, proceed now, he argued, as follows: (1) calculate the $\frac{1}{2}n(n-1)$ possible values of $s_d^2 = \sum_i (d_{uv\,i} - \bar{d}_{uv})^2/m$; (2) multiply each by a factor $m/(m-1)$ so that its

* One variety would appear twice in this mean and its yield must be suitably weighted.

† Gosset at a later date made comments on this point and on the assumption involved in getting a pooled estimate of standard errors that might differ; see (**11**, pp. 285 and 282).

expectation becomes σ^2; (3) sum these quantities and divide by their number. Thus the final estimate of σ^2 becomes

$$s^2 = \frac{2 \sum\limits_{u,v} \sum\limits_{i} (d_{uv\cdot i} - \bar{d}_{uv})^2}{n(n-1)(m-1)}. \qquad \text{......(14)}$$

As I shall explain later, this is exactly the estimate which would now be used, only it would be calculated in a more direct manner. The division of Beaven's plots into sets of eight which I have shown in Fig. 1, would to-day be termed a division into blocks (though the blocks are not similar in shape), and the arrangement of the different varieties within a block would be called balanced rather than random. Thus already in 1912 Beaven and Gosset together had gone a long way towards reaching one form of the present-day experimental technique.

Having obtained the estimate s^2 of (14), Gosset was then able to consider the significance of the difference between any pair of varieties by calculating the ratio

$$x = \frac{\bar{d}_{uv}\sqrt{m}}{s}, \qquad \text{......(15)}$$

and referring to Sheppard's tables.* His method was to place the eight varieties in order of magnitude of the character under consideration and, by applying the test as a foot-rule to selected differences, draw reasoned conclusions as to the existence or absence of real variety differences. A test (R. A. Fisher's z-test) which would determine whether as a whole the eight variety means differed significantly would clearly have been useful, but sound common sense could make the difference test yield reliable results.

This method was applied to the English and Irish chess-board results; the computation was lengthy and many pages of a large notebook of Gosset's are filled with the calculations. G. U. Yule carried out the Cambridge computations in consultation with Gosset. But, however laborious the work, the conclusions obtained from the analysis combined with results of large scale tests played an important part in securing the steady improvement that was being effected in the quality of Irish grown barley.

It is perhaps of historical interest to note a more general formula that Gosset was using at this time to obtain a common estimate of standard deviation from data classified into a number of groups with possibly different means.† The formula would not now be regarded as satisfactory, but it illustrates well the slow progress of the human mind to its final goal.

Suppose that N observations of a variable x are divided into n groups of unequal size, that x_{ti} is the ith observation in the tth group; further that m_t is

* The common estimate, s^2, of (14) is based on so many observations that Gosset probably had not considered whether $\bar{d}_{uv}/s$ could be referred to the z-distribution.

† I have taken the expression from a letter of 1912 to Beaven.

the number and $\bar{x}_t$ the mean in that group. Then Gosset took as an estimate of a supposed common within-group variance, σ^2, the expression

$$s^2 = \frac{1}{N} \sum_{t=1}^{n} \frac{m_t}{m_t - 1} \sum_{i=1}^{m_t} (x_{ti} - \bar{x}_t)^2. \qquad \ldots\ldots(16)$$

Since the expectation of $\sum_i (x_{ti} - \bar{x}_t)^2$ is $(m_t - 1)\sigma^2$ and $N = \sum_t m_t$ it will be seen that the expectation of s^2 is σ^2. Except in the case where m_t is the same for every group, which was the case he was concerned with in the chess-board analysis, the factors weighting the sums of squares are not, however, those which we now know give an estimate of σ^2 having minimum sampling error. When however $m_t = m$ his estimate assumed the correct form

$$s^2 = \frac{1}{n(m-1)} \sum_t \sum_i (x_{ti} - \bar{x}_t)^2. \qquad \ldots\ldots(17)$$

Had he applied formula (16) to the chess-board problem in a case where the number of plots was not the same for all varieties, his final estimate would have been less satisfactory.

During the war period of 1914–19 the analysis of the chess-board results was discontinued. In 1920 Gosset took over responsibility for the statistical aspects of the barley experiments conducted at a number of centres by the Irish Department of Agriculture, and this made him particularly interested in the possibilities of Beaven's new half drill strip method of arrangement. Correspondence with Beaven is full of discussion of the possibilities of this method and of the best way of analysing the results. At the same time he was in touch with R. A. Fisher who was beginning to turn his great mathematical powers to similar problems at Rothamsted.

The next reference I can find to the chess-board analysis is early in 1923, when Beaven had asked Gosset to explain again the procedure he had used ten years before. The final lap of the long passage to an "analysis of variance" is of sufficient historical and personal interest to place on record. On 29 March 1923 Gosset writes.

> I enclose a note on the chess-board error. I was using the formula before the war and see no reason to repent of it. I am writing Fisher asking him to look it over and if necessary criticize.

The method given is that which I have described above, involving the calculation of the $\frac{1}{2}n(n-1)$ squares of differences. It was naturally a lengthy procedure, and I find a brief note of Beaven on the papers, after working through an example: "Conclusion (if any possible) from above is that P.E. with chess-boards might be guessed at almost as well as calculated." It needed a "Student" with his facility for doing calculations in spare moments on the back of an envelope to cope with such computations. But the author of the method himself was not

content and on 9 April in the second half of a letter started on the 6th, he writes again to Beaven:

Since writing the above I have had a vision on the subject of chess-board error and enclose a rough proof of my new method. I have written to Yule asking him whether he is in fact working at chess-board error and enclosing a similar proof. If he is *not* I shall be inclined to write it up and shall ask your leave to use the No. 1 chess-board of 1913 as an illustration. If he *is*, he has doubtless got something as good or better, and he can put mine in the W.P.B.

To use my new method with 15 plots, each of 8 varieties (1) find the square of the S.D. of the whole 120 plots, Σ^2; (2) after calculating the averages of the eight varieties, find the square of the S.D. of these eight figures, σ_8^2; (3) after calculating the averages of the fifteen groups of eight, find the square of the S.D. of these fifteen figures, σ_{15}^2. Then the P.E. of the error of a comparison should be

$$0{\cdot}6745\sqrt{\frac{2\times 8(\Sigma^2-\sigma_8^2-\sigma_{15}^2)}{120-8-15}}.*$$

In calculating the S.D.'s do not use the $(n-1)$ divisor.

The "rough proof" of the method which he enclosed was as follows: it will be seen to be on similar lines to that given in the paper "On testing varieties of cereals" (11, pp. 282–3) except for the omission of the term $-\sigma_e^2/mn$ referred to in the published paper, which resulted in a divisor of $mn-m-n$ instead of $mn-m-n+1$.

Memorandum

Let m plots of each of n varieties be chessboarded. There will be m groups each containing one of each of the n varieties. If Σ^2 be the variance of the nm plots, it may be considered to be composed of three parts which as a first approximation may be taken as uncorrelated:

(1) The real differences between the varieties, σ_v^2,

(2) The errors common to each group of n, σ_c^2,

(3) The remaining casual errors, σ_e^2.

Of these the last is the only part that affects the comparison of varieties since the differences which we intend to measure compose (1), and (2) is eliminated by the process of chessboarding.

It remains to find the best estimate of (1), (2) and (3) given Σ^2, the averages of the n varieties, and those of the m groups.

Now if σ_n be the S.D. of the averages of the n varieties

$$\sigma_n^2=\sigma_v^2+\sigma_e^2/m,\dagger$$

and if σ_m be the S.D. of the averages of the m groups

$$\sigma_m^2=\sigma_c^2+\sigma_e^2/n.$$

Also
$$\Sigma^2=\sigma_v^2+\sigma_c^2+\sigma_e^2.$$

* This is the P.E. of the difference between two means of fifteen plots. It must be squared and multiplied by $m=15$ to get into the form of (18) below. [E. S. P.]

† The expression on the right-hand side should have been $\sigma_v^2+\sigma_e^2(n-1)/mn$, this is equal to the expectation of σ_n^2. Similar corrections to the σ_e^2 term are required in the next two equations. [E. S. P.]

Hence $$\Sigma^2-\sigma_n^2-\sigma_m^2=\sigma_e^2\left(1-\frac{1}{n}-\frac{1}{m}\right),$$

therefore $$\sigma_e^2=\frac{mn(\Sigma^2-\sigma_n^2-\sigma_m^2)}{mn-m-n}.$$

Whence the others follow, and the error of a comparison between a pair of varieties is

$$\sqrt{\frac{2}{m}}\,\sigma_e=\sqrt{\frac{2n(\Sigma^2-\sigma_n^2-\sigma_m^2)}{mn-m-n}}.$$

In the next letter to Beaven of 20 April Gosset writes:

Now as to chess-board error. About a week after I sent the proposed simplified method to you and Yule, I got a note from Fisher via Somerfield giving the same method in rather more technical language. Next I got a reply from Yule saying that the method was new and giving it his blessing more or less, and finally I got a p.c. from Fisher this morning saying that the divisor should be $mn-m-n+1$ not $mn-m-n$. Anyhow the thing seems to have some weight behind it now.

It should give the same result as my original method....

That the agreement between the two results depends on the identity*

$$\frac{2\sum\limits_{u,v}\sum\limits_{i}(d_{uv\cdot i}-\bar{d}_{uv})^2}{n(n-1)\,(m-1)}=2\times\frac{\sum\limits_{u}\sum\limits_{i}(x_{ui}-\bar{x})^2-m\sum\limits_{u}(\bar{x}_u-\bar{x})^2-n\sum\limits_{i}(\bar{x}_i-\bar{x})^2}{(n-1)\,(m-1)}\quad\text{......(18)}$$

was shown by Fisher in the letter Gosset quotes in the footnote to p. 283 of his paper (11). The expression on the left-hand side is taken from formula (14) above, while that on the right represents the estimate of the sampling variance of the difference between two single plot yields obtained by the usual analysis of variance method.

Fisher's application of the method was given in a joint paper with W. A. Mackenzie on "Studies in crop variation", received by the *Journal of Agricultural Science* on 20 March 1923 and published in July. The theory was illustrated on an experiment with potatoes "planted in triplicate on the 'chess-board' system"; the arrangement of the plots was not so well balanced as in Beaven's chess-board and as yet no question of randomization was considered. The paper contained what was I think the first published arrangement of numerical data in an analysis of variance table (then described as analysis of variation), and a method was given of testing for the significance of the treatment (or variety) sum of squares, taken as a whole.

"Student's" paper (11) was read before the Society of Biometricians and Mathematical Statisticians on 28 May 1923 and published in *Biometrika* in the following December. In obtaining the formula of the memorandum even with the slip which no doubt he would later have found out himself, and in the description of the method of procedure given to Beaven, he had so evidently after long searching reached the essential conception of breaking up a total sum

* In this notation $d_{uv\cdot i}=x_{ui}-x_{vi}$ or is the difference between uth and vth varieties in the ith block. $\bar{x}_u$, $\bar{x}_i$ and $\bar{x}$ are the variety, the block and the grand mean respectively. There are n varieties and m blocks.

of squares into parts* that I feel his achievement should be put on record. As we have seen, in his modest way he was ready to have his results thrown into the waste paper basket, if another statistician could improve on his work! Whether his mathematics could ever have shown unaided that if no variety differences existed:

(1) the expressions $\Sigma^2 - \sigma_n^2 - \sigma_m^2$ and σ_n^2 of his memorandum were independent,

(2) were each distributed in a modified form of the distribution he had discovered in 1908,

(3) gave a ratio whose distribution law was a Pearson type VI curve;

all this is doubtful. But, as he would have said himself, why speculate, these further results were derived by Fisher; the problem was therefore solved and a new chapter opened.

The 1923 paper (**11**) contains much else of interest besides this handling of the chess-board type of experiment. It starts with an historical survey of the development of experiments aiming at the comparison of cereals and concludes with a critical discussion of the half drill strip method. The simple theme which I have referred to on many occasions runs through the whole and takes form in a final concluding sentence:

It is shown that methods (2) [chess-board] and (3) [half drill strip] depend for their accuracy on the fact that the nearer two plots of ground are situated, the more highly are the yields correlated, so that we are able to increase the effect of the last term of the equation

$$\sigma_{A-B}^2 = \sigma_A^2 + \sigma_B^2 - 2r_{AB}\sigma_A\sigma_B$$

(where A and B are the varieties to be compared) by placing the plots to be compared with one another as near together as possible.

Later papers

In his later papers Gosset tended to avoid, as far as possible, the introduction of mathematics and he would ask his friends to regard him as a non-mathematician. Thus he forwarded his paper on the Lanarkshire milk experiment (**17**) to Karl Pearson with the words:

I hope you will find it interesting, though its chief merit to the likes of me (that there are no mathematics in it), will hardly commend it to you.

Or again, writing to me in 1926 regarding the original χ^2 paper (Karl Pearson, 1900) he remarked:

I have now read the χ^2 paper in *Phil. Mag.* **50**. It may be divided into three parts, one that I can follow as a man who could cut a block of wood into the rough shape of a boat with his penknife might appreciate a model yacht cut and rigged to scale, the second I can

* His original approach to statistics through Airy's book made this a natural way of regarding things; see the formula (5) I have quoted above. There are points in Gosset's proof in (**11**, p. 282) also reminiscent of Airy, *Theory of Errors of Observations* (1875, p. 46).

only compare to a conjuring trick of which I haven't got the key (such for mexaple as tho transformation to polar co-ordinates on p. 158) and lastly quite a small part which I think I can understand.

When at last, after the war, an increasing number of men trained as mathematicians began to turn their attention to statistics, it was not perhaps surprising that one whose mathematical training had ceased with Oxford Mods. in the nineties should refuse to regard himself as a mathematician. Besides, the increasing responsibilities of his work as a brewer left him little time or inclination to follow out in detail the continuous elaboration of the theory of mathematical statistics. As a result, in his relatively rare publications he tended to concentrate on simple exposition of the function of statistical method. The best examples of such work are:

(1) The paper on "Errors of routine analysis" of 1927 (**15**) which develops more fully a theme he had touched on before (**4** and **8**), and shows how some recent theoretical work on the distribution of "range" in small samples might be made to give a useful working tool for the analyst.

(2) Two admirable papers on the use of statistical methods in agriculture, both unfortunately rather inaccessible to the ordinary student: "Mathematics and Agronomy", 1926 (**14**), and the article on "Yield Trials" in *Baillière's Encyclopedia of Scientific Agriculture* 1931 (**16**).

This recession from the mathematical approach of his earlier papers had other consequences. In the first place it meant that during a period of rapid advance in statistical technique there was available, for almost anyone in need of advice, a statistician of great practical experience and unusual insight, whom the inquirer could be sure would not be carried away by the fascination of any mathematical model into allowing abstract theory to step beyond its proper sphere. On the other hand there were certain disadvantages; Gosset's avoidance of a mathematical statement of his case sometimes, as in his last two papers (**21**), (**22**), made it difficult for others to grasp an idea or method which probably was clear enough in his own mind. The theory of probability is based on mathematics, and beyond a certain point there are dangers in introducing it into practice without a precise mathematical statement of the assumptions underlying the method of procedure.

If we return to 1923, it is clear that Gosset welcomed with enthusiasm the new methods that R. A. Fisher was developing. The neatness of the arrangement of calculations in an analysis of variance table for example, appealed to him. It brought to the rather laborious calculation methods of his own a simplification whose value he was quick to realize. The introduction of t as the ratio of a deviation to an estimate of its standard error, in place of his own criterion z, and the use of degrees of freedom, appealed to him at once because of the greater generality; as a result he calculated extended values of the probability integral of t to replace his old z tables and published these in 1925 (**13**) in conjunction

with a theoretical contribution of Fisher's. In print and in correspondence he emphasized the importance of randomness. "The experiments", he wrote in 1926 (**14**, p. 711), "must be capable of being considered to be a *random* sample of the population to which the conclusions are to be applied. Neglect of this rule has led to the estimate of the value of statistics which is expressed in the crescendo 'lies, damned lies, statistics'."

This paper of 1926 contains perhaps the extreme limit to which he ventured in allowing the toss of a coin or a die to decide the arrangement of plots in an agricultural experiment. On the last page (p. 719) he suggests the arrangement of four varieties in an 8×8 square, in which two plots of each variety are to fall in each row and each column. Subject to this restriction the arrangement was to be obtained in a random manner.

He must soon, however, have realized the disadvantages of such a procedure. If A, B, C and D represent the varieties, a possible if unlikely run of luck might lead to the following pattern of plots in one corner of the square:

A	*A*	*C*	*D*	*C*	
A	*A*	*C*	*D*	*B*	
B	*C*	*A*	*A*	*D*	
C	*C*	*C*	*B*	*A*	

Fig. 2.

Should this chance juxtaposition of many A-plots happen to coincide with a "fertility summit" or "depression" in the field, the resulting statistical analysis of plot yields might easily attribute a characteristic to the variety A which it did not possess. His practical mind could not accept such a state of affairs. To know in advance that if an experiment was carried out with a particular pattern of plots there was quite a chance that it would be misleading, and to continue with this pattern—this was a course he was not prepared to follow. It was no compensation to be told that in the long run, if the verdict of the random toss was accepted and the 5% significance level of mathematical tables used in the statistical analysis, then misleading results would be obtained only 5 times in 100. In his own words (**22**, p. 366):

> It is of course perfectly true that *in the long run*, taking all possible arrangements, exactly as many misleading conclusions will be drawn as are allowed for in the tables, and anyone prepared to spend a blameless life in repeating an experiment would doubtless confirm this; nevertheless it would be pedantic to continue with an arrangement of plots known beforehand to be likely to lead to a misleading conclusion.

His withdrawal from the out and out randomization position is illustrated in his article of 1931 on "Yield Trials" (16). Here he speaks of the Latin square arrangement as ideal in the types of experiment for which it is suited, because it combines the elements of balance and randomness, but he is critical of the randomized block arrangement because of the risk involved of getting misleading results. He gives the following illustration of a balanced or equalized block design which he had recommended to a horticultural correspondent, comparing ten treatments with five replications:

G	*H*	*E*	*C*	*A*	Block I
F	*D*	*J*	*B*	*I*	
H	*J*	*D*	*F*	*E*	Block II
B	*G*	*I*	*A*	*C*	
E	*I*	*A*	*G*	*D*	Block III
J	*B*	*C*	*H*	*F*	
C	*F*	*B*	*I*	*J*	Block IV
A	*E*	*G*	*D*	*H*	
D	*A*	*F*	*J*	*B*	Block V
I	*C*	*H*	*E*	*G*	

Fig. 3.

In this example the assignment of treatments to plots in Block I is random, but each successive block has its arrangement more and more controlled, so that (i) each of the five columns contains one plot only of the ten varieties, (ii) *A*, *D*, *E*, *F* and *J* occur in the top row of their block three times and in the lower row twice, while for *B*, *C*, *G*, *H* and *I* the position is reversed, an arrangement as nearly balanced as possible for an odd number of blocks.

In advocating the introduction of this element of balance, he did not consider that the random element could be dispensed with; but he believed that if a regular pattern was used to equalize the more probable variations in fertility there were still sufficient complications to leave the residual variations random enough to justify from the practical point of view the application of probability theory. It was here that he disagreed and was eventually forced into open controversy with R. A. Fisher and the Rothamsted school.

This is not the place to enter into detail regarding the nature of this controversy, which resulted in Gosset's last paper published a few months after his

death (22). It is however well to emphasize that his attitude was closely related to the type of agricultural problem with which he had had most experience, the development of improved strains of barley. In such a case as this he saw that success was only likely to result from a comparison of two or more strains in a number of years and in a number of different localities. Small scale investigations must be followed by others in which the technique conformed as far as possible to ordinary agricultural practice. In each case some experimental plan was needed which would give the yields, let us say, of variety A and variety B on the experimental area with as little error as possible, that is to say freed from bias such as might be introduced by changes in fertility, patches of weed, etc. Provided that the error of the difference (yield of A−yield of B) could be kept low, he was satisfied with a knowledge of its probable upper limit and did not mind if he was told that the ratio of this difference to the estimate of its standard error in a particular experiment could not be referred with mathematical precision to a table of probabilities. He was interested primarily in the behaviour of the difference from farm to farm and year to year, and experience had shown him, beyond any possibility of doubt, that small scale balanced plot experiments followed by larger scale tests with the half drill strip method of Beaven's, the purpose of which any intelligent farmer could understand, had achieved remarkable success in the improvement of barley. If it were argued that fully randomized experimental designs would have achieved the same or better results he would not have denied this dogmatically, but he felt doubtful on the point because his perusal of reports on such experiments showed to his mind an unduly high proportion of inconclusive results. He would also have added that with the staff, the ground and other facilities available in the investigations for whose planning he was responsible, fully randomized designs could not have been carried out. This was his attitude in writing the Statistical Society paper of 1936 (21).

In his final paper (22) he attacked his critics on their own ground by pointing out that in the experiment at a single station a balanced arrangement of plots in blocks was on the whole more likely to detect variety differences than a random arrangement when those differences were really large and therefore important, although for small differences the reverse would be true.

The ultimate decision on these points can hardly be expected as yet; it will come in time, perhaps after 10 or after 20 years, when there has been ample opportunity for the practical experimenter, freed from the weight of authority, from fear of mathematics on the one side and from the fascination of a new technique on the other, to judge from accumulated experience what methods have been most worth while having regard to the results they have achieved.

In addition to these papers on agricultural subjects a brief reference may be given to some other published work of the last few years:

(1) A paper on "The Lanarkshire milk experiment", 1931 (17); his suggestion

that the experiment should be repeated on a more precise but far less expensive scale by using pairs of twins involves a characteristic introduction of his paired difference plan.

(2) Two papers on certain implications of F. L. Winter's selection experiments with maize 1933 (19) and 1934 (20). The plant breeder's problem of improving varieties of cereal by continued selection had long been of interest to him in connexion with barley and in these papers he discusses the bearing of these experimental results upon evolutionary theory.

(3) A number of short but suggestive contributions to the discussion of papers read before the Industrial and Agricultural Research Section of the Royal Statistical Society (see references on p. 249 below).

Extracts from letters

I have spoken more than once of Gosset's correspondence; the professional statistician, whether he be attached to a university or research station, receives and expects to receive appeals for advice which will continue to increase through life as his circle of contacts grows. But with Gosset the position was somewhat different; to provide advice to correspondents all over the world was in no way part of his job. Yet he gave that help unstintingly and unless it could be described as brewery business, he gave it out of his own time. Advice as to how to plan a particular experiment, or explanations of misunderstood points in statistical theory, while of extreme value at the time to the individual who receives them are rarely of interest to the general reader. Nevertheless, I believe that a few quotations from letters will add to the record of Gosset's personality by showing something of his patience, his practical mind, his suggestiveness and his characteristic freedom of expression.

The first quotations are taken from a long letter written to me in 1926. At that time I had been trying to discover some principle beyond that of practical expediency which would justify the use of "Student's" ratio $z=(\bar{x}-m)/s$ in testing the hypothesis that the mean of the sampled population was at m. Gosset's reply had a tremendous influence on the direction of my subsequent work, for the first paragraph contains the germ of that idea which has formed the basis of all the later joint researches of Neyman and myself. It is the simple suggestion that the only valid reason for rejecting a statistical hypothesis is that some alternative hypothesis explains the observed events with a greater degree of probability. The second part of the letter probably put into my mind the very extensive plan of sampling from non-normal populations which we carried out in the Department of Statistics at University College during the next few years.

Letter I

From a letter of W. S. G. to E. S. P., dated 11 May 1926.

In your large samples with a known normal distribution you are able to find the chance that the mean of a random sample will lie at any given distance from the mean of the population. (Personally I am inclined to think your cases are best considered as mine taken to the limit n large.) That doesn't in itself necessarily prove that the sample is not drawn randomly from the population even if the chance is very small, say ·00001: what it does is to show that if there is any alternative hypothesis which will explain the occurrence of the sample with a more reasonable probability, say ·05 (such as that it belongs to a different population or that the sample wasn't random or whatever will do the trick) you will be very much more inclined to consider that the original hypothesis is not true.

I can conceive of circumstances, such for example as dealing a hand of 13 trumps after careful shuffling *by myself*, in which almost any degree of improbability would fail to shake my belief in the hypothesis that the sample was in fact a reasonably random one from a given population.

*   *   *   *   *   *   *

I'm more troubled really by the assumption of normality and have tried from time to time to see what happens with other population distributions, but I understand that you get correlation between s and m with *any* other population distribution.

Still I wish you'd tell me what happens with the even chance population ▭ or such a one as △: it's beyond my analysis.

*   *   *   *   *   *   *

If Student is wrong it is up to you to give us something better. You see one must experiment and frequently it is quite out of the question, from considerations of cost or of impossibility of duplicating conditions in the time scale, to do enough repetitions to define one's variability as accurately as one could wish. It's no good saying "Oh these small samples can't prove anything". Demonstrably small samples *have* proved all sorts of things and it is really a question of defining the amount of dependence that can be placed on their results as accurately as we can. Obviously we lose by having a poor definition of the variability but *how much do we lose?*

Letter II, with its enclosure which, for reasons I have forgotten, was never published, was written shortly after K. P. had made an editorial comment on "Sophister's"* (1928) interpretation of the distribution of "Student's" ratio in samples from a non-normal population. It had been found that in such cases the distribution of t was asymmetrical, but that the distribution of $|t|$ (or of t^2) followed very closely the standard normal-theory form, i.e. if the distribution of t was curtailed on one side of the origin this was balanced by a corresponding extension on the other side. The letter also refers to a suggestion of bringing up at a meeting of the International Statistical Institute the question of differentiating between the symbols used for probable error and standard error.

* "Sophister" like "Mathetes" was the *nom de plume* of a disciple of "Student". The particular sampling investigation in question had been sketched out by Gosset and myself before "Sophister" came to spend a year in the Biometric Laboratory.

Letter II

Holly House,
Blackrock,
Co. Dublin.
May 18th, '29.

Dear Pearson,

I was rather amused to see your letter open with an apology for delay in writing as I have for some time been acutely conscious that I have been in arrears. However, last things first.

(i) I agree that Z's second suggestion though sound is not workable. Your idea of raising the question at Warsaw seems to me to suggest the right way of getting to work. I think they should raise the question on the grounds (*a*) that $\pm$ is being used in two senses, (*b*) that the prob. error is no longer the slightest use to anyone and (*c*) that as the tables are in terms of the S.D. a simple notation such as : or ; or anything of the sort is required.

(ii) I fancy you give me credit for being a more systematic sort of cove than I really am in the matter of limits of significance. What would actually happen would be that I should make out P_t (normal) and say to myself "that would be about 50 : 1; pretty good but as it may not be normal we'd best not be too certain", or "100 : 1; even allowing that it may not be normal it seems good enough" and whether one would be content with that or would require further work would depend on the importance of the conclusion and the difficulty of obtaining suitable experience.

One so often finds that the importance (and even occasionally the direction of the result) of varying one factor, change from experiment (or experience) to experiment according to the accompanying variations in other factors, that it often doesn't pay to make too certain of any one result.

E.g. You may have two varieties of barley one of which will give the best yield in one season or place while the other will win in another season or place; hence we have to sample places and seasons widely rather than aim at being meticulously accurate at all places sampled: there must be economy of effort.

* * * * * * *

Lastly I am enclosing a short note in reply to the Editorial footnote. Probably you are going to say all that is at all useful in it in your next paper, and in any case I haven't the least intention of indulging in a controversy, so suppress it unless you think it will clear up our position. All the same I think it is a pity to let the thing go by default without any comment.

Yours v. sincerely,
W. S. Gosset.

Suggested Note for "Biometrika"

17th May, 1929.

In his footnote on page 422 of Sophister's paper the Editor asks, "Supposing 50 per cent of prisoners tried for murder were acquitted and the remainder found guilty should we be right *in the long run* to drop the trial and toss up for judgment?" This, if I may say so, is hardly what Sophister proposes to do. If I may deal first with the Editorial analogy the position is rather, "The evidence before the court is such that the chances are even that the prisoner committed the murder". Doubtless if more evidence were forthcoming we should know more about it; as it is, an English Court will acquit, though the inexorable Justice of Shan Tien would condemn the prisoner to piecemeal slicing, unless of course sufficiently weighty evidence for the defence could be imperceptibly introduced within the Mandarin's sleeve. But, seriously, a better illustration can be drawn from the practice of

Insurance where in the first place the premium is calculated on the Healthy Male table and, I suppose, originally this was the only basis after a medical examination. But the material which supplied the experience for the H.M. table can be subdivided into various classes, by professions and occupations, by stature or eye colour, total abstainers or moderate drinkers and so forth, which further investigation may find to have expectations of life which do not accord with the table. The life expectation of some of these classes is probably taken into consideration by the Companies—I doubt whether a Lion Tamer, however healthy, could insure at the ordinary rates—but no company, as it well might, charges a lower rate of premium for the descendants of centenarians or a higher for orphans; they are most unfairly lumped together just as Sophister proposes to do with his samples from unknown populations. In effect he says, "This small sample is from an unknown population, which *may* be normal; it probably is not far from normal; if it *is* normal we use the table justly, if it is anormal but symmetrical we can still use the table with sufficient accuracy; even if it is skew, about which we cannot be sure—much less about the direction of the skewness—we shall in the long run draw much the same proportion of correct inferences as if it were normal." Admittedly our ignorance of the nature of the population introduces an element of uncertainty which no sensible person will ignore when using the tables, but recent work, and not least Sophister's, shows that this uncertainty, while not altogether negligible, is much less than we had any right to expect.

Student.

The suggestion in Letter III of 1932 ultimately led to the production of tables of percentage limits of the ratio of (*a*) range in a sample of n observations to (*b*) an independent estimate of standard deviation, which are to be published shortly in *Biometrika*. From the beginning of his analysis of the results of the chess-board experiments, Gosset had wondered how best to judge what differences among variety means were significant. While the ratio of (*a*) the difference between any two means selected at random to (*b*) the estimate of standard error could be referred to "Student's" distribution or, if desired, the significance of the set as a whole could be judged by Fisher's z-test, it was not possible to treat selected differences in either of these ways. In the article in *Baillière's Encyclopedia* (**16**, p. 1358) he refers to a method suggested by Fisher of taking the differences between individual variety yields and the mean yield. He felt however that a knowledge of the probability levels of "studentized" range would in addition be very useful; on this could be based a rough test of the kind he had suggested in his paper on "Errors of routine analysis" (**15**, p. 161).

Letter III

St. James's Gate,
Dublin.
Jan. 29th, '32.

Dear Pearson,

Many thanks for your letter and enclosure: as I am at the moment

"The Cook and the Captain bold
And the mate of the Nancy brig",

I have handed all the lot to Mathetes till such time as I can get a chance of dealing with it which should be sometime next week.

I have been meaning to write to you for some time *re* the proposals for the use of range and sub-range which I made in my last letter to you. Of course there is a serious crab which

I had at one time recognised and then forgotten in that the thing would have to be "Studentised": the only measure of the S.D. is provided by a limited number of degrees of freedom. Whether one could get an approximate correction for this with moderately small numbers by reducing still further the degrees of freedom or whether it would be necessary as Fisher suggested when I mentioned the matter to him (he was here lecturing) to dive into the depths of hyperspace to produce the jewel I am not clear, but obviously something would have to be done about it.

* * * * * * *

Yrs. v. sincerely
W. S. Gosset.

Letters IV and V of 1936, which Dr Beaven has kindly allowed me to reproduce, deal with the interpretation of the results of half drill strip barley experiments carried out at six stations in England; the two varieties compared were Plumage Archer and Beaven's 35/7. The second letter followed a reply from Beaven discussing the position in terms of betting on two horses, whose form varies on different courses. The argument illustrates Gosset's outlook on the function of large scale experiments to which I have already referred.

Letter IV

From a letter of W. S. G. to E. S. B., dated 8 January 1936.

If you derive the S.E. from a set of 10 strips at one station, you are sampling "comparisons between plots grown at a certain station in the weather of 1935" and can draw the appropriate conclusion, e.g. that at Sprouston it is quite certain that Beaven's 35/7 would have beaten Plumage Archer in any sound arrangement of plots in 1935.

When however you regard the six stations as a small sample of the barley land of England you can very nearly draw the conclusion that Beaven's 35/7 would on the average have beaten Plumage Archer if compared all over the barley land of England in 1935.

The chance that so favourable a result would have happened if there were really no difference between them is only 1/38, i.e. the odds are 37 to 1 against it's happening. This is very nearly significant but as you know, what odds are to be considered significant is a matter of convention—or taste.

Naturally, in calculating the S.E. (not really an *error* at all) of the second conception where the variation from Station to Station depends as much, (or much more . . . than), on the differential response to weather and soil as on the soil errors taken account of in each station, one takes no particular account of the S.E.'s at the individual stations: one merely rejoices because the Half Drill Strip method has largely eliminated the errors due to soil position and left us mainly the differential response aforesaid, which would have affected the result to a greater or less extent in every field of barley-growing England and which we have assumed that we have sampled by the six results which we have examined.

I hope I have made the distinction clear between the S.E. of the result at one station, which is rightly derived from the plots grown at that station but which only enables us to judge whether the result is significant for that station, and the S.E. of the whole series, derivable only from the six mean results of the six stations but which enables us to make an estimate of the result of comparing the barleys "everywhere", where "everywhere" represents the whole extent of country that may properly be considered to be sampled by the six stations.

Letter V

Davan Hollow,
Denham,
Bucks.
14. 1. 36.

Dear Beaven,

I don't think your analogy is quite exact: this is mine.

The two horses 35/7 and P.A. are known to vary somewhat from day to day and also to be very much affected by the particular course on which they are running.

They have raced ten times at Sprouston and 35/7 has won every time by amounts varying from one furlong to two furlongs. At Sprouston then you may lay longish odds on 35/7. At Cambridge they raced ten times and on the average 35/7 won by 50 yds, the amounts varying from 270 yds in favour of 35/7 to 170 in favour of P.A. You would not therefore bet very heavily on 35/7 at Cambridge. At four other places 35/7 beat P.A. on average by various amounts. What odds is to be given on another hitherto untried course?

You are surely as much influenced by the narrowness of the margin at Cambridge as by the width of it at Sprouston: the new course may resemble the one with just as much likelihood as the other and may even as far as you can see favour P.A. rather than 35/7, since your knowledge of the difference between courses rests on only six cases.

Furthermore a new method of training may reduce the variation so that the Sprouston results may lie between $1\frac{1}{4}$ and $1\frac{3}{4}$ furlongs and the Cambridge between 160 yds in favour of 35/7 and 60 yds in favour of P.A., without altering very much* the odds on a series of races on a new course, since the chief source of variation remains the reaction of the horses to the courses and not the day to day variation which alone is measured by the variation on a single course.

*   *   *   *   *   *   *

Yours v. sincerely
W. S. Gosset.

* But since the smaller day to day variation prevents an accidentally high or low value of mean obscuring the real value of the course there is a better chance of getting the right odds—not of getting higher odds.

Letter VI was written at the time when Gosset was putting together his last paper (**22**).

Letter VI

Dart Cottage,
Postbridge,
Devon.
19. iv. 37.

Dear Pearson,

Many thanks for yours of 10th; I feel I'm rather wasting your time but as long as you ask questions you must expect to get answers. You have given my reason for not changing the level of significance *viz.* that while balancing certainly *tends* to produce a lower real error and consequently higher calculated error one cannot say how much one has succeeded in any particular case. I therefore content myself with pointing out that the tendency is beneficial, not only are the cases missed of comparatively little value but one actually gets more conclusions of real value.

*   *   *   *   *   *   *

Now I was talking about Cooperative experiments and obviously the important thing in such is to have a low real error, not to have a "significant" result at a particular station. The latter seems to me to be nearly valueless in itself. Even when experiments are carried out only at a single station, if they are not mere five finger exercises, they will have to be

part of a series in time so as to sample weather and the significance of a single experiment is of little value compared with the significance of the series—which depends on the real error not that calculated for each experiment.

But in fact experiments at a single station *are* almost valueless; you can say "In heavy soils like Rabbitsbury potatoes cannot utilise potash manures", but when you are asked "What *are* heavy soils like Rabbitsbury?" you have to admit—until you have tried elsewhere—that what you mean is "At Rabbitsbury etc." And that, according to X may mean only "In the old cow field at Rabbitsbury". What you really want to find out is "In what soil and under what conditions of weather do potatoes utilise the addition of potash manures?"

To do that you must try it out at a representative sample of the farms of the country and correlate with the characters of the soil and weather. It may be that you have an easy problem, like our barleys which come out in much the same order wherever—in reason—you grow them or like Crowther's cotton which benefitted very appreciably from nitro-chalk in seven stations out of eight, but even then what you really want is a low real error. You want to be able to say not only "We have significant evidence that if farmers in general do this they will make money by it", but also "we have found it so in nineteen cases out of twenty and we are finding out why it doesn't work in the twentieth". To do that you have to be as sure as possible which *is* the 20th—your real error must be small.

* * * * * * *

Tedin:* Somerfield sent me the number and I have just had time to glance at it. T. put down three kinds of patterns of Latin Squares (5×5) on various uniformity trials. There were

Two Knight's moves:

A	*B*	*C*	*D*	*E*
D	*E*	*A*	*B*	*C*
B	*C*	*D*	*E*	*A*
E	*A*	*B*	*C*	*D*
C	*D*	*E*	*A*	*B*

Two Diagonals:

A	*B*	*C*	*D*	*E*
E	*A*	*B*	*C*	*D*
D	*E*	*A*	*B*	*C*
C	*D*	*E*	*A*	*B*
B	*C*	*D*	*E*	*A*

and a number of randoms.

Of course all Latin squares are "balanced" but one wouldn't care too much for the "Diagonal" arrangement and the Knight's move would, I think, be preferred to all others. In conformity with this Tedin found a slight tendency for the Knight's move to give a low actual and a high calculated error while the diagonal tends to give a high actual and a low calculated error. The whole thing is not worth worrying about but is interesting as an illustration of what actually happens when we depart from artificial randomisation: I would Knight's move every time!

Yours
W. S. G.

P.S. Beaven after all got some slight ailment which prevented his being in the chair for Bartlett's paper: I proposed the vote of thanks....I was heard without enthusiasm but there were no cat calls!

Such are my impressions of Gosset and of his work. Others will have different views on the relative importance of his many contributions to statistics; on his rightness or wrongness. The experimentalist will have seen him in a different light from the mathematician; his personal friends will have realized aspects of his character which his correspondents could not see. But all who have known him will agree that he possessed almost more of the characteristics of the perfect

* A reference to the paper by O. Tedin (1931).

statistician than any man of his time. They will agree, too, on the essential balance and tolerance of his outlook, and on that something which a friend of his schooldays has described as an "immovable foundation of niceness" which made him through life the same friendly dependable person, quiet and unassuming, who worked not for the making of personal reputation, but because he felt a job wanted doing and was therefore worth doing well.

BIBLIOGRAPHY OF "STUDENT'S" PAPERS

(1) 1907. "On the error of counting with a haemacytometer." *Biometrika*, **5**, 351.

(2) 1908. "The probable error of a mean." *Biometrika*, **6**, 1.

(3) 1908. "Probable error of a correlation coefficient." *Biometrika*, **6**, 302.

(4) 1909. "The distribution of the means of samples which are not drawn at random." *Biometrika*, **7**, 210.

(5) 1911. Appendix to paper by W. B. Mercer and A. D. Hall on "The experimental error of field trials." *J. Agric. Sci.* **4**, 128.

(6) 1913. "The correction to be made to the correlation ratio for grouping." *Biometrika*, **9**, 316.

(7) 1914. "The elimination of spurious correlation due to position in time or space." *Biometrika*, **10**, 179.

(8) 1917. "Tables for estimating the probability that the mean of a unique sample of observations lies between $-\infty$ and any given distance of the mean of the population from which the sample is drawn." *Biometrika*, **11**, 179.

(9) 1919. "An explanation of deviations from Poisson's law in practice." *Biometrika*, **12**, 211.

(10) 1921. "An experimental determination of the probable error of Dr Spearman's correlation coefficients." *Biometrika*, **13**, 263.

(11) 1923. "On testing varieties of cereals." *Biometrika*, **15**, 271.

(12) 1924. Note by "Student" with regard to his paper "On testing varieties of cereals." *Biometrika*, **16**, 411.

(13) 1925. "New tables for testing the significance of observations." *Metron*, **5**, 105.

(14) 1926. "Mathematics and Agronomy." *J. Amer. Soc. Agron.* **18**, 703.

(15) 1927. "Errors of routine analysis." *Biometrika*, **19**, 151.

(16) 1931. Article on "Yield Trials" in *Baillière's Encyclopedia of Scientific Agriculture*. **2**, 1342.

(17) 1931. "The Lanarkshire milk experiment." *Biometrika*, **23**, 398.

(18) 1931. "On the 'z' test." *Biometrika*, **23**, 407.

(19) 1933. "Evolution by selection. The implications of Winter's selection experiment." *Eugen. Rev.* **24**, 293.

(20) 1934. "A calculation of the minimum number of genes in Winter's selection experiment." *Ann. Eugen., Lond.*, **6**, 77.

(21) 1936. "Co-operation in large-scale experiments." A discussion opened by W. S. Gosset. *J.R. Statist. Soc.* Suppl. **3**, 115.

(22) 1937. "Random and balanced arrangements." *Biometrika*, **29**, 363.

A FEW SHORTER CONTRIBUTIONS

(*a*) *Letters to "Nature"*.

29 November 1930, **126**, 843: "Agricultural Field Experiments".

14 March 1931, **127**, 404: "Agricultural Field Experiments".

5 December 1936, **138**, 971: "The half drill strip system of Agricultural Experiments".

(*b*) ***Contributions to discussions at meetings of the Industrial and Agricultural Research Section of the Royal Statistical Society.***

J.R. Statist. Soc. Suppl. (1934), **1**, 18; (1936), **3**, 173; (1937), **4**, 89, 170.

REFERENCES TO PAPERS BY OTHER AUTHORS

CAVE, F. E. (1904). *Proc. Roy. Soc.* **74**, 403.
HELMERT, F. R. (1876). *Astr. Nachr.* **88**, S. 122.
HOOKER, R. H. (1905). *J.R. Statist. Soc.* **68**, 696.
MACDONELL, W. R. (1901). *Biometrika*, **1**, 219.
MERCER, W. B. & HALL, A. D. (1911). *J. Agric. Sci.* **4**, 107.
NEYMAN, J. (1937). *Lectures and conferences on mathematical statistics.* Graduate School of U.S. Dept. of Agriculture, Washington.
PEARSON, K. (1900). *Phil. Mag.* **50**, 157.
PEARSON, K. (1907). *Drapers' Company Research Memoirs.* Biometric Series, **4**.
PEARSON, K. (1913). *Biometrika*, **9**, 116.
SOPER, H. E. (1913). *Biometrika*, **9**, 91.
"SOPHISTER" (1928). *Biometrika*, **20***a*, 389.
TEDIN, O. (1931). *J. Agric. Sci.* **21**, 191.
WHITTAKER, L. (1914). *Biometrika*, **10**, 36.
WOOD, T. B. & STRATTON, F. J. M. (1910). *J. Agric. Sci.* **3**, 417.

THE DISTRIBUTION OF SPEARMAN'S COEFFICIENT OF RANK CORRELATION IN A UNIVERSE IN WHICH ALL RANKINGS OCCUR AN EQUAL NUMBER OF TIMES

By M. G. KENDALL, SHEILA F. H KENDALL AND B. BABINGTON SMITH

PART I. THEORETICAL DETERMINATION OF THE SAMPLING DISTRIBUTION OF SPEARMAN'S COEFFICIENT OF RANK CORRELATION

Introduction

1. If n individuals are ranked according to two qualities in the orders $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$, where the X's and the Y's are permutations of the numbers 1 to n, the coefficient of rank correlation between the rankings is defined as

$$\rho = 1 - \frac{6 \overset{n}{\underset{i=1}{S}} (d_i^2)}{n^3 - n}, \qquad \ldots\ldots(1)$$

where $d_i = X_i - Y_i$. The coefficient ρ, introduced by Spearman (1904), is the product-moment coefficient of correlation between X and Y.

If and only if the correspondence between the two rankings is perfect, i.e. $X_i = Y_i$, $\rho = 1$. On the other hand, if and only if the two rankings are exactly inverted, i.e. $X_i = Y_{n-i+1}$, $\rho = -1$. In other cases ρ lies between these limits.

2. In order to judge of the significance of a value of ρ it is necessary to consider the distribution of values obtained by correlating an arbitrary order, which may conveniently be taken as the order 1, 2, ..., n, with all other permutations of the numbers 1 to n. In practice it is generally more convenient to consider the distribution of the quantity $S(d^2)$, which is related to ρ by equation (1).

3. Certain simple properties of this distribution are obtainable immediately.

(a) Any value of $S(d^2)$ must be even. For $S(d) = 0$, being the difference of the sums of the first n natural numbers; hence the number of odd values of d is even, and so is the number of odd values of d^2.

(b) The possible values of $S(d^2)$ range from 0 to $\frac{1}{3}(n^3 - n)$ and hence there are $\frac{1}{6}(n^3 - n) + 1$ of them.

(c) The distribution is symmetrical, about a central value if $\frac{1}{6}(n^3 - n)$ is even, or about two adjacent central values if $\frac{1}{6}(n^3 - n)$ is odd. This follows from the fact that to any given value of ρ corresponding to a permutation P there will correspond a negative value of ρ of the same absolute value arising from P inverted.

For, if the permutation P is $X_1, X_2, \ldots, X_n$ the inverted permutation is $X_n, X_{n-1}, \ldots, X_1$. $S(d^2)$ calculated from P is then $S(X_i - i)^2$ and that from P inverted is $S(X_i - \overline{n+1} + i)^2$. The sum of these two is

$$S(X_i^2) + S(i^2) - 2S(X_i i) + S(X_i^2) + S(n+1-i)^2 - 2S\{X_i(n+1-i)\}.$$

The first, second, fourth and fifth items in this expression are each equal to the sum of the squares of the first n natural numbers; the third and sixth, taken together, are equal to $-2(n+1)\,S(X_i) = -2(n+1)\,.\,\frac{1}{2}n(n+1)$. Hence the sum of the two values of $S(d^2)$

$$= \tfrac{4}{6}n(n+1)(2n+1) - n(n+1)^2$$
$$= \tfrac{1}{3}(n^3-n).$$

The result follows simply from equation (1).

(*d*) It follows from (c) that all odd moments about the mean of the distribution of $S(d^2)$ vanish.

4. A further important result, due to "Student", was given by Karl Pearson (1907), namely that the second moment of ρ is

$$\mu_2(\rho) = \frac{1}{n-1}, \qquad \ldots\ldots(2)$$

from which it follows at once that

$$\mu_2\{S(d^2)\} = \frac{n^2(n+1)^2(n-1)}{36}. \qquad \ldots\ldots(3)$$

5. The distribution of ρ has recently been considered by Hotelling and Pabst (1936), who have proved the remarkable theorem that as n tends to infinity the distribution tends to normality.

The distributions for low values of n, so far as they have been obtained, deviate quite considerably from normality and it has not previously been made clear how great n must be for normality to be assumed with much confidence, particularly in the determination of significance levels. Unfortunately ρ is mainly of service in the range $n=10$ to $n=30$, i.e. precisely where the doubt lies. It is the aim of the present paper to throw some light on this crepuscular territory.

Expression for the distribution of $S(d^2)$

6. Consider the deviations between the order 1, 2, ..., n and an order X. If one deviation is known, then certain deviations become impossible for other ranks. For instance, if the deviation d_1 between X_1 and 1 is $(n-1)$, then $X_1=n$, and it is impossible for the deviation between X_2 and 2 to be $(n-2)$; or for the deviation between X_3 and 3 to be $(n-3)$, and so on. Consider then the array:

$$\begin{array}{ccccccc}
n-1 & n-2 & n-3 & \ldots & 2 & 1 & 0 \\
n-2 & n-3 & n-4 & \ldots & 1 & 0 & -1 \\
n-3 & n-4 & n-5 & \ldots & 0 & -1 & -2 \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
2 & 1 & 0 & \ldots & -(n-5) & -(n-4) & -(n-3) \\
1 & 0 & -1 & \ldots & -(n-4) & -(n-3) & -(n-2) \\
0 & -1 & -2 & \ldots & -(n-3) & -(n-2) & -(n-1)
\end{array}$$

If d_k has the value in the rth row and the kth column then d_l cannot have the value in the rth row and the lth column; and so on.

In fact, any permissible set of deviations is given by taking n entries from the above table so that no row or column contributes more than one entry.

Hence to get $S(d^2)$ for any permissible set write

$$E = \begin{Bmatrix} a^0 & a^1 & a^4 & a^9 & \dots & a^{(n-1)^2} \\ a^1 & a^0 & a^1 & a^4 & \dots & a^{(n-2)^2} \\ a^4 & a^1 & a^0 & a^1 & \dots & a^{(n-3)^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a^{(n-1)^2} & a^{(n-2)^2} & a^{(n-3)^2} & a^{(n-4)^2} & \dots & a^0 \end{Bmatrix}, \qquad \dots\dots(4)$$

and $S(d^2)$ is given by the index of a of one of the terms obtained from E by choosing n factors so that no row or column appears more than once and multiplying them together. Thus the distribution of $S(d^2)$ is given by the totality of $n!$ terms which can be constructed in that way. E will be taken to be equal to the polynomial in a given by the sum of these terms.

7. E bears an obvious analogy to the determinant, but it cannot be regarded as such and expanded accordingly. If it could the distribution of $S(d^2)$ would be obtained without difficulty, for a determinant with the elements of E as given above may be shown to be equal to

$$(1-a^2)^{n-1}(1-a^4)^{n-2}(1-a^6)^{n-4}\dots(1-a^{2(n-1)}). \qquad \dots\dots(5)$$

E, in fact, lacks the fundamental property of the determinant in that it does not change sign if two rows or columns are interchanged.

8. Nevertheless certain of the rules of determinantal algebra remain true for E The most valuable is that E may be expanded in terms of its minors of any order in the usual way. Expansion of this type is, in fact, rather easier with E than with the determinant, for all terms of E are essentially positive and there are no difficulties with signs. We have used this expansion repeatedly in obtaining the distributions given below. There are also certain devices which assist the expansion of E in virtue of its symmetry. Two which have been found useful are as follows:

(*a*) Any minor of E is symmetrical in powers of a, i.e. is of the form

$$A_0a^k + A_2a^{k-2} + A_4a^{k-4} + \dots + A_4a^{m-4} + A_2a^{m-2} + A_0a^m.$$

(*b*) The effect of shifting a minor bodily across E is to multiply each term of its expansion by a constant power of a.

This property may be proved thus: Let an r-rowed minor be

$$M = \begin{Bmatrix} a^{(\alpha-\kappa_1)^2} & a^{(\beta-\kappa_1)^2} & a^{(\gamma-\kappa_1)^2} & \dots & a^{(\zeta-\kappa_1)^2} \\ a^{(\alpha-\kappa_2)^2} & a^{(\beta-\kappa_2)^2} & a^{(\gamma-\kappa_2)^2} & \dots & a^{(\zeta-\kappa_2)^2} \\ \dots & \dots & \dots & \dots & \dots \\ a^{(\alpha-\kappa_r)^2} & a^{(\beta-\kappa_r)^2} & a^{(\gamma-\kappa_r)^2} & \dots & a^{(\zeta-\kappa_r)^2} \end{Bmatrix}.$$

If we shift the minor λ places to the left we have

$$M' = \begin{Bmatrix} a^{(\alpha-\kappa_1-\lambda)^2} & a^{(\beta-\kappa_1-\lambda)^2} & a^{(\gamma-\kappa_1-\lambda)^2} & \dots & a^{(\zeta-\kappa_1-\lambda)^2} \\ a^{(\alpha-\kappa_2-\lambda)^2} & a^{(\beta-\kappa_2-\lambda)^2} & a^{(\gamma-\kappa_2-\lambda)^2} & \dots & a^{(\zeta-\kappa_2-\lambda)^2} \\ \dots & \dots & \dots & \dots & \dots \\ a^{(\alpha-\kappa_r-\lambda)^2} & a^{(\beta-\kappa_r-\lambda)^2} & a^{(\gamma-\kappa_r-\lambda)^2} & \dots & a^{(\zeta-\kappa_r-\lambda)^2} \end{Bmatrix}$$

The factor a^{λ^2} in the first row is common to all terms and may thus be brought outside the curly bracket. Similarly for the other rows. We shall then be left with items of type $a^{(\alpha-\kappa_1)^2-2\lambda(\alpha-\kappa_1)}$. The factor $a^{2\lambda\kappa_1}$ is common to all members of the first row and thus may be brought outside the bracket. The factor $a^{-2\lambda\alpha}$ is common to all terms of the same column and may also be brought outside. Proceeding thus with similar terms we shall have

$$M' = Ma^{r\lambda^2-2\lambda S(\alpha)+2\lambda S(\kappa)},$$

which is the result stated.

For example the minors

$$M = \begin{Bmatrix} a^0 & a^1 & a^4 \\ a^1 & a^0 & a^1 \\ a^4 & a^1 & a^0 \end{Bmatrix} = a^0 + 2a^2 + 2a^4 + a^6$$

and

$$M' = \begin{Bmatrix} a^4 & a^9 & a^{16} \\ a^1 & a^4 & a^9 \\ a^0 & a^1 & a^4 \end{Bmatrix} = a^{12}(a^0 + 2a^2 + 2a^4 + a^6)$$

are related by

$$M' = Ma^{12}.$$

9. Even with these aids the evaluation of E is a tedious business, though straightforward enough. We have found it for values of n from 1 to 8, the resulting distributions of $S(d^2)$ being given in Table I.

As checks on the resulting distribution, it will be remembered that the total frequency is $n!$ and the second moment about the mean $\dfrac{n^2(n+1)^2(n-1)}{36}$.

10. Additional checks on the lower values of $S(d^2)$ may be obtained by considering directly the number of permutations giving $S(d^2) = 0, 2, 4, \dots,$ etc. For example, with n ranks, a value of 2 can only arise by the sum of terms $1+1$, which in turn can only arise by the interchange of two adjacent terms in the order $1, 2, \dots, n$. This number of values is therefore $(n-1)$. A value of $S(d^2)$ equal to 4 can only arise as $1+1+1+1$, i.e. by two interchanges of pairs of adjacent terms, and the number of ways such interchange can be made is $\dfrac{(n-2)!}{2!(n-4)!}$. Expressions

TABLE I

Distributions of $S(d^2)$ for values of n from 1 to 8

Values of n

$S(d^2)$	1	2	3	4	5	6	7	8
0	1	1	1	1	1	1	1	1
2	.	1	2	3	4	5	6	7
4	.	.	0	1	3	6	10	15
6	.	.	2	4	6	9	14	22
8	.	.	1	2	7	16	29	47
10	.	.	.	2	6	12	26	54
12	.	.	.	2	4	14	35	70
14	.	.	.	4	10	24	46	94
16	.	.	.	1	6	20	55	129
18	.	.	.	3	10	21	54	124
20	.	.	.	1	6	23	74	178
22	.	.	.	.	10	28	70	183
24	.	.	.	.	6	24	84	237
26	.	.	.	.	10	34	90	238
28	.	.	.	.	4	20	78	276
30	.	.	.	.	6	32	90	264
32	.	.	.	.	7	42	129	379
34	.		.	.	6	29	106	349
36	.	.	.	.	3	29	123	380
38	.	.	.	.	4	42	134	400
40	.	.	.	.	1	32	147	517
42	.	.	.	.	.	20	98	394
44	.	.	.	.	.	34	168	542
46	.	.	.	.	.	24	130	492
48	.	.	.	.	.	28	175	640
50	.	.	.	.	.	23	144	557
52	.	.	.	.		21	168	666
54	.	.	.		.	20	144	595
56	.	.	.	.		24	184 (median)	776
58	.	.	.	.	.	14	.	684
60	.	.	.	.	.	12	.	786
62	.	.	.	.	.	16	.	718
64	.	.	.	.	.	9	.	922
66		.	.	.	.	6	.	745
68	.	.	.	.	.	5	.	917
70	.	.	.	.	.	1	.	781
72	.	.	.	.		.	.	982
74	.	.	.	.	.	.	.	826
76	.	.	.	.	.	.	.	950
78	.	.	.	.	.	.	.	844
80	.	.	.	.	.	.	.	1066
82	.	.	.	.	.	.	.	845
84	.	.	.	.	.	.	.	936 (median)
Total	1	2	6	24	120	720	5040*	40320*

* Total of whole distribution, only the median value and the values on one side of the median being shown in this table.

of this type, however, rapidly become very complicated. For values of $S(d^2)$ up to and including 22 we find the following frequencies:

$S(d^2)$	Frequency
0	1
2	$n-1$
4	$\binom{n-2}{2}$
6	$\binom{n-3}{3}+2\binom{n-2}{1}$
8	$\binom{n-4}{4}+4\binom{n-3}{2}+\binom{n-2}{1}$
10	$\binom{n-5}{5}+6\binom{n-4}{3}+2\binom{n-3}{2}+2\binom{n-3}{1}$
12	$\binom{n-6}{6}+8\binom{n-5}{4}+3\binom{n-4}{3}+4\binom{n-4}{2}+4\binom{n-4}{2}+2\binom{n-3}{1}$
14	$\binom{n-7}{7}+10\binom{n-6}{5}+4\binom{n-5}{4}+12\binom{n-5}{3}+6\binom{n-5}{3}$ $+4\binom{n-4}{2}+2\binom{n-4}{1}+4\binom{n-3}{1}+4\binom{n-4}{2}$
16	$\binom{n-8}{8}+12\binom{n-7}{6}+24\binom{n-6}{4}+5\binom{n-6}{5}+\binom{n-4}{2}$ $+12\binom{n-5}{3}+8\binom{n-6}{4}+8\binom{n-5}{2}+6\binom{n-5}{3}$ $+8\binom{n-4}{2}+\binom{n-3}{1}+4\binom{n-5}{2}+4\binom{n-4}{1}$
18	$\binom{n-9}{9}+14\binom{n-8}{7}+40\binom{n-7}{5}+8\binom{n-6}{3}+6\binom{n-7}{6}$ $+3\binom{n-5}{3}+24\binom{n-6}{4}+10\binom{n-7}{5}+24\binom{n-6}{3}+4\binom{n-5}{2}$ $+8\binom{n-6}{4}+8\binom{n-5}{2}+12\binom{n-5}{3}+2\binom{n-4}{2}+3\binom{n-3}{1}$ $+6\binom{n-6}{3}+8\binom{n-5}{2}+4\binom{n-4}{1}+2\binom{n-5}{1}$

$S(d^2)$ Frequency

20

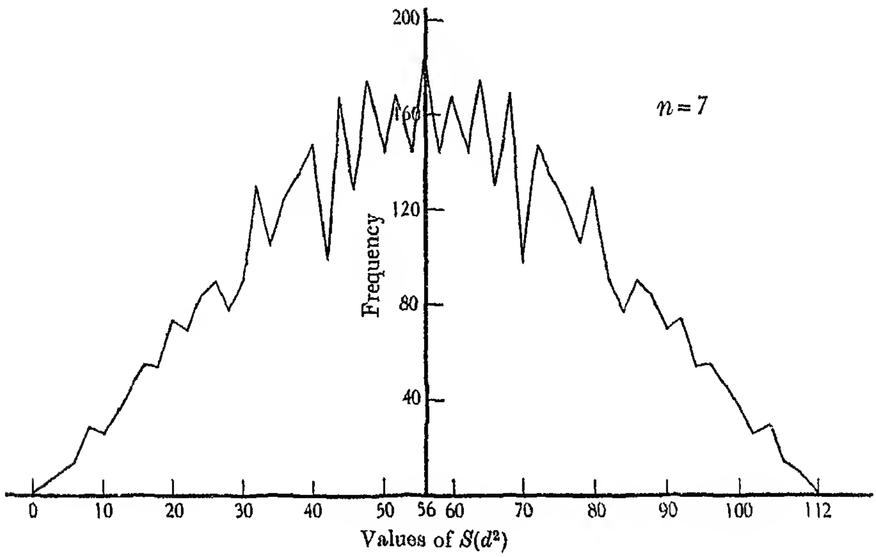
$$\binom{n-10}{10}+16\binom{n-9}{8}+60\binom{n-8}{6}+32\binom{n-7}{4}+7\binom{n-8}{7}$$
$$+6\binom{n-6}{4}+40\binom{n-7}{5}+12\binom{n-6}{3}+12\binom{n-8}{6}+4\binom{n-6}{2}$$
$$+48\binom{n-7}{4}+12\binom{n-6}{3}+10\binom{n-7}{5}+24\binom{n-6}{3}+4\binom{n-5}{2}$$
$$+16\binom{n-6}{4}+16\binom{n-5}{2}+3\binom{n-5}{3}+6\binom{n-4}{2}+\binom{n-3}{1}$$
$$+8\binom{n-7}{4}+8\binom{n-6}{2}+12\binom{n-6}{3}+8\binom{n-5}{2}+4\binom{n-4}{1}$$
$$+4\binom{n-6}{2}+6\binom{n-5}{1}$$

22

$$\binom{n-11}{11}+18\binom{n-10}{9}+84\binom{n-9}{7}+80\binom{n-8}{5}+8\binom{n-9}{8}$$
$$+10\binom{n-7}{5}+60\binom{n-8}{6}+48\binom{n-7}{4}+6\binom{n-6}{3}+14\binom{n-9}{7}$$
$$+12\binom{n-7}{3}+80\binom{n-8}{5}+24\binom{n-7}{3}+24\binom{n-7}{4}+12\binom{n-8}{6}$$
$$+48\binom{n-7}{4}+12\binom{n-6}{3}+8\binom{n-6}{2}+20\binom{n-7}{5}+48\binom{n-6}{3}$$
$$+8\binom{n-5}{2}+4\binom{n-6}{4}+4\binom{n-5}{2}+9\binom{n-5}{3}+2\binom{n-4}{2}$$
$$+10\binom{n-8}{5}+24\binom{n-7}{3}+4\binom{n-6}{2}+16\binom{n-7}{4}+16\binom{n-6}{2}$$
$$+12\binom{n-6}{3}+8\binom{n-5}{2}+10\binom{n-4}{1}+6\binom{n-7}{3}+12\binom{n-6}{2}$$
$$+6\binom{n-5}{1}+2\binom{n-6}{1} \qquad \ldots\ldots(6)$$

These results (the first four of which were given by Hotelling & Pabst (1936)) can, of course, be written more simply, but are set out in the above form so that the method of obtaining them may be followed more easily. Each term corresponds to a different type of arrangement required to give the specified value of $S(d^2)$. The successive values of the frequency do not appear to conform to any simple law, and it is not to be expected that they should, inasmuch as the terms composing them depend on the partitions of even numbers into squares.

11. The distributions of Table I are peculiar in several respects. For lower values of n they are distinctly bimodal. For $n = 7$ and $n = 8$ the frequency polygons



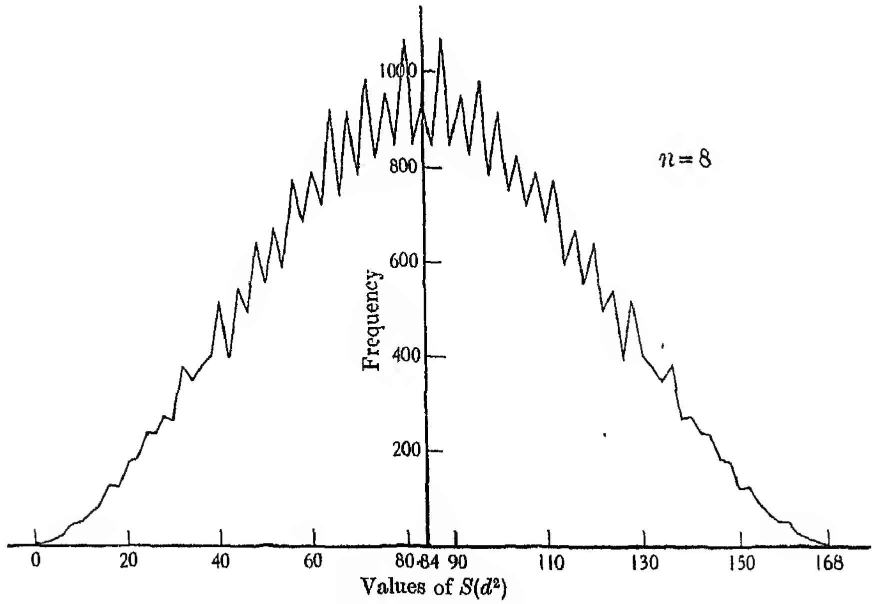


Fig. 1. Frequency polygons of $S(d^2)$ for $n = 7$ and $n = 8$.

have an unusual serrated profile, as may be seen from Fig. 1, though normality is beginning to emerge.

12. The value of μ_4 for the distribution of ρ was given by Hotelling & Pabst (1936) as

$$\mu_4 = \frac{3(25n^4 - 13n^3 - 73n^2 + 37n + 72)}{25n(n+1)^2 (n-1)^3}, \qquad \text{......(7)}$$

and it follows that

$$\beta_2 = 3 + \frac{24}{100} \frac{36 - 5n - 19n^2}{n^3 - n}. \qquad \text{......(8)}$$

The values of $\beta_2 - 3$ for certain values of n are shown in Table II. The distribution is platykurtic, approaching mesokurtosis as n becomes larger. But as might have been expected β_2 fails to reveal the serrated appearance of the frequency polygon for low n.

TABLE II

Values of $\beta_2 - 3$ in the distribution of $S(d^2)$ for various values of n

n	$\beta_2 - 3$	n	$\beta_2 - 3$
1	—	15	−0·308
2	−2·000	20	−0·230
5	−0·928	25	−0·184
10	−0·464	30	−0·153

13. So far as we have calculated the distribution, the serrations in the frequency polygon show no signs of disappearing over the main range, and it is not immediately obvious what happens as n becomes larger and the polygon tends to normality. From the form of Fig. 1, however, it would seem that the tails of the curve smooth out first, and that the smoothness runs up towards the apex of the distribution as n tends to infinity.

PITMAN'S APPROXIMATION

14. It will be clear, we think, that at least for $n = 8$ or less the normal curve offers only an indifferent representation of the distribution of $S(d^2)$. For example, in the distribution for $n = 8$, the chance of getting a value of $S(d^2)$ outside the range 14–154 (i.e. as great as or greater than 156, or as small as or less than 12) is 0·0107 (the nearest point to a 1 % significance level in this discontinuous case). If the distribution were taken to be normal with the same mean and standard deviation (in this case $12\sqrt{7}$) the chance would be 0·0233. A correction for continuity would not improve matters materially.

15. Pitman (1937), observing that the first four moments of ρ are approximately the same as those of the B-distribution

$$df = \frac{1}{B(\frac{1}{2}, \frac{1}{2}n - 1)} (1 - x^2)^{\frac{n}{2} - 2} dx, \qquad \text{......(9)}$$

has suggested that the probability integral of this curve, i.e. a Pearson Type II curve, may be used for that of ρ; and says that the true values agree well with

those of the approximate distribution even for values of n as low as 6, which is apparently the greatest value of n for which he had the actual distribution. This is true over the greater part of the range, and the B-distribution appears to give a fair idea of the true values of the significance points.

For instance, with $n=8$ the distribution becomes

$$df = \frac{1}{B(\frac{1}{2}, 3)}(1-x^2)^2\,dx,$$

and by direct integration the probability of a value greater than x in absolute value is

$$1-\frac{15}{8}\left(x-\frac{2x^3}{3}+\frac{x^5}{5}\right). \qquad \text{......(10)}$$

The chance of getting a value of $S(d^2)$ outside the range 14–154 is, as above, 0·0107. The chance calculated from the formula (10), with a correction for continuity,* is 0·0098.

Similarly, the chance of getting a value outside the range 26–142 is 0·0576. That given by the formula (10) is 0·0561.

16. It may be expected that for larger values of n Pitman's approximation is closer, and would probably provide a satisfactory test of significance for practical purposes.

Furthermore, there is an extremely close relation between ρ and another measure of rank correlation suggested elsewhere (Kendall, 1938) the sampling distribution of which may be readily obtained. (This relation is discussed below in Part 3 of this paper.) For these reasons we have not thought it necessary to embark on the labour of determining E for values of n greater than 8.

It is, perhaps, worth noting that the probability integral of the curves (9) may be related to "Student's" t-integral by the transformation

$$t = x\sqrt{(n-2)}/\sqrt{(1-x^2)},$$

which gives, on substitution in (9),

$$df = \frac{1}{\sqrt{(n-2)}\,B(\frac{1}{2}, \frac{1}{2}n-1)} \frac{dt}{\left(1+\frac{t^2}{n-2}\right)^{\frac{n-1}{2}}},$$

the "Student" form with $n-2$ degrees of freedom.

The deviate x corresponding to a value S of $S(d^2)$, with continuity corrections, is

$$\frac{S}{\frac{1}{6}(n^3-n)+1}-1$$

If n is large the denominator term may be taken to be $\frac{1}{6}(n^3-n)$, and to this approximation

$$x = -\rho.$$

* The continuity correction was made by assuming the range of the B-curve to be equivalent to the range -1 to $\frac{1}{3}(n^3-n)+1$ for $S(d^2)$, i.e. the terminal frequencies were assumed distributed over a range of two units, one unit on each side of the terminal ordinates.

We thus reach the notable result that the approximate significance points of ρ may be determined from "Student's" t-distribution with $n-2$ degrees of freedom by writing

$$t = \rho\sqrt{(n-2)}/\sqrt{(1-\rho^2)}.$$

A transformation of the same kind may be used to test the significance of a value of the product-moment coefficient of correlation of a sample of n from an uncorrelated bivariate normal universe. The resemblance between such a coefficient and ρ becomes even more striking when it is remembered that the former has the same variance as has ρ in the case under consideration.

PART II. EXPERIMENTAL DISTRIBUTIONS OF ρ

17. As an alternative to calculating E for values of n greater than 8 we have conducted some experiments to find empirically the distribution of $S(d^2)$ for $n=10$ and $n=20$.

For the cases $n=10$ and $n=20$ sets of permutations of the numbers 0 to 9 and 1 to 20 were constructed from the tables of Tippett (1927) in the manner described below. There is reason to suppose that the coefficients or values of $S(d^2)$ calculated from these data are a random and representative selection from the possible values.

Method of obtaining data for permutations of 10

18. The 2000 permutations of the numbers 0 to 9 were obtained in the following way: the observer went through Tippett's numbers (beginning on the first page and reading across), writing down the digits as they occurred but omitting those which had occurred already in the particular permutation he was constructing. When nine out of the possible ten had occurred the tenth was filled in without reference to the tables and the observer began on a new permutation. Thus, the first 51 of Tippett's numbers are

2952 6641 3992 9792 7979 5911 3170 5624

4167 9524 1545 1396 720

The first permutation will be 2 9 5 6 4 1 3 7 0 8, involving the first twenty-eight numbers. The last figure contributed from the table is 0, the 8 being filled in automatically; so beginning with the twenty-ninth figure, we find that the second permutation is 5 6 2 4 1 7 9 3 0 8.

19. The 2000 permutations were found to require 39,183 digits, an average of 19·59 per permutation, and hence cover practically the whole of Tippett's table. We discuss the relationship between the expected and observed average run in the Appendix. So far as our tests show, the values of $S(d^2)$ obtained may

be regarded as reasonably random, and are representative of the theoretical distribution within sampling limits.*

Method of obtaining data for permutations of 20

20. For the permutations of 20 a rather different technique was adopted. Each pair of Tippett's digits was taken to give a number from 1 to 20, numbers of 21 or greater being reduced by subtracting multiples of 20. Thus, the numbers 21, 41, 61, 81 were taken as giving the number 1, the numbers 00, 40, 60, 80 as giving the number 20, and so on. This process can be carried out at sight, and, as for the permutations of 10, an observer went through Tippett's tables reading out each pair so obtained. The first eight Tippett's digits, 2952 6641, thus yield four numbers 9, 12, 6, 1.

In order to eliminate errors, a second observer was provided with a working sheet of paper and a slip of cardboard on which were written the numbers 1, 2, ..., 20 in their natural order. This was adjusted on the working sheet so as to lie a little higher up the sheet than the twenty spaces in which the random permutation was to be written, the number 1 lying above the first space and so on. The numbers read from Tippett's tables were then numbered serially on the working sheet in the order in which they occurred. Thus, for the sequence 9, 12, 6, 1, the second observer would write the numbers 1, 2, 3, 4 on the working sheet at the places indicated by the figures 9, 12, 6, 1 on the cardboard slip. When the first observer read out a number which had occurred already in the permutation under construction, the second observer ignored it—and could do so without possibility of error because the space allotted to that number had already been filled. As before, when nineteen numbers had been obtained, the last was filled in automatically.

21. The above process does not give the permutation as it occurs in the table, but a second permutation which has elsewhere been called the conjugate of the first (Kendall, 1938). Thus, to take a simple case, consider the order

$$\begin{array}{cccccc} A & 1 & 2 & 3 & 4 & 5 \\ B & 4 & 3 & 2 & 5 & 1 \end{array}$$

Rearrange B in the order 1, 2, 3, 4, 5, and rearrange A in the same manner, so that any A-number, which lies above a B-number in the above, continues to do so, thus

$$\begin{array}{cccccc} A' & 5 & 3 & 2 & 1 & 4 \\ B' & 1 & 2 & 3 & 4 & 5 \end{array}$$

If we repeat the process on A' and B' we get back to A and B. B and A' may be called conjugate permutations.

* The fact that the permutations emanate from Tippett's numbers would no doubt be accepted by many as sufficient guarantee that the resulting values of $S(d^2)$ are a random sample. We ourselves felt that further tests were necessary, for reasons given at length elsewhere (Kendall & Babington Smith, 1938).

It is easy to see that if a permutation occurs in Tippett's tables as B the procedure described above will result in A' being written down.

22. If B is a random permutation A' will also be a random permutation. Perhaps of more importance for present purposes is the fact that the coefficient ρ between the order 1, 2, . ., n and a permutation B is the same as between 1, 2, ..., n and the conjugate permutation A'. For ρ depends only on the differences d, and these are the same in the case A, B as in the case A', B', though occurring in a different order Hence, for the purposes of calculating, either the permutation or its conjugate may be used. The choice between them is entirely a matter of convenience, and, as has already been stated, we find that writing down the conjugate is simpler and far less liable to error.

Considerations of space prevent us from giving these random permutations in full, but we should be glad to place them at the disposal of any workers who could find use for them. They can, of course, be used to construct permutations of objects fewer in number than 10 or 20 as the case may be, by the omission of certain numbers.

DISTRIBUTION OF $S(d^2)$ IN THE RANKINGS OF 10

23. The distribution of values of $S(d^2)$ in the 2000 rankings of 10 is given in Table III.

As the frequencies in individual compartments are rather small we have grouped them in Table IV.

It is evident at once that the distribution as judged by the first moment about the universe mean ($S(d^2)=165$) is sufficiently symmetrical. In fact the first moment for the grouped distribution of Table IV is 0·18 (expected value zero). The variance of the theoretical distribution, from equation (3), is 3025, and hence the standard error of the mean of 2000 sets is 1·23. The observed deviation from expectation is thus well within sampling limits.

The same is true of the second moment, the observed value for the grouped data of Table IV being 2980·7 (expected value 3025), deviation $-44{\cdot}3$. The standard error of the second moment $=\mu_2\sqrt{\{(\beta_2-1)/n\}} = 84$ approximately.

24. So far as these tests go, therefore, the distribution conforms to expectation. Notable features of the grouped distribution of Table IV are the antimodes at $S(d^2) = 136$–144 and $S(d^2) = 206$–214. It would appear that for $n=10$ a certain amount of irregularity still persists and that the assumption of normality cannot be confidently made near the mean. More important from the sampling point of view is the behaviour of the distribution near the tails. Even for a sample as large as 2000, the frequencies occurring in the ends of the range are hardly big enough to allow a reliable comparison to be made with the theoretical frequencies given by the B-curves. Comparisons for some broad groupings, however, indicate a reasonable concordance. For example the B-curve for $n=10$

TABLE III

Distribution of 2000 values of $S(d^2)$ for $n=10$

$S(d^2)$	Frequency			$S(d^2)$	Frequency		
	1st thousand	2nd thousand	Total		1st thousand	2nd thousand	Total
0	—	—	—	330	—	—	—
2	—	—	—	328	—	—	—
4	—	—	—	326	—	—	—
6	—	—	—	324	—	—	—
8	—	—	—	322	—	—	—
10	—	—	—	320	—	—	—
12	—	—	—	318	—	—	—
14	—	—	—	316	—	—	—
16	—	—	—	314	—	—	—
18	—	—	—	312	—	—	—
20	—	—	—	310	—	—	—
22	—	—	—	308	1	—	1
24	—	—	—	306	2	—	2
26	—	1	1	304	—	1	1
28	—	1	1	302	—	—	—
30	—	—	—	300	1	1	2
32	—	—	—	298	—	—	—
34	—	—	—	296	1	1	2
36	—	1	1	294	1	—	1
38	1	—	1	292	1	—	1
40	2	—	2	290	1	2	3
42	—	2	2	288	2	—	2
44	1	2	3	286	1	—	1
46	1	2	3	284	1	—	1
48	3	3	6	282	6	2	8
50	2	—	2	280	1	2	3
52	3	3	6	278	2	1	3
54	1	2	3	276	1	2	3
56	1	3	4	274	2	3	5
58	2	1	3	272	4	2	6
60	3	2	5	270	5	3	8
62	2	1	3	268	1	4	5
64	6	4	10	266	4	9	13
66	3	3	6	264	1	5	6
68	4	5	9	262	1	4	5
70	2	4	6	260	8	3	11
72	2	2	4	258	3	3	6
74	4	3	7	256	4	4	8
76	6	3	9	254	4	7	11
78	9	4	13	252	5	7	12
80	9	5	14	250	8	7	15
82	5	5	10	248	5	3	8
84	10	4	14	246	3	7	10

TABLE III

Continued

$S(d^2)$	Frequency			$S(d^2)$	Frequency		
	1st thousand	2nd thousand	Total		1st thousand	2nd thousand	Total
86	3	7	10	244	3	4	7
88	7	5	12	242	7	3	10
90	10	5	15	240	5	2	7
92	4	8	12	238	6	5	11
94	7	7	14	236	7	2	9
96	4	7	11	234	13	5	18
98	11	7	18	232	9	5	14
100	10	4	14	230	5	13	18
102	14	13	27	228	9	7	16
104	11	9	20	226	8	9	17
106	10	5	15	224	8	7	15
108	9	11	20	222	11	9	20
110	13	10	23	220	9	11	20
112	8	18	26	218	12	9	21
114	11	9	20	216	13	13	26
116	16	12	28	214	8	4	12
118	11	15	26	212	6	10	16
120	12	12	24	210	13	14	27
122	8	5	13	208	12	6	18
124	14	5	19	206	7	12	19
126	9	12	21	204	8	16	24
128	12	14	26	202	14	8	22
130	7	17	24	200	12	12	24
132	8	10	18	198	9	16	25
134	15	13	28	196	11	14	25
136	6	10	16	194	13	14	27
138	9	8	17	192	12	12	24
140	10	10	20	190	13	13	26
142	15	9	24	188	16	16	32
144	5	14	19	186	12	18	30
146	14	17	31	184	12	14	26
148	9	15	24	182	13	11	24
150	13	11	24	180	15	11	26
152	8	10	18	178	16	20	36
154	14	14	28	176	8	14	22
156	14	7	21	174	15	16	31
158	20	8	28	172	11	18	29
160	12	6	18	170	13	14	27
162	10	14	24	168	14	10	24
164	17	21	38	166	10	20	30
Totals	502	480	982	Totals	498	520	1018

TABLE IV

Distribution of the 2000 sets of Table III, condensed

$S(d^2)$ (inclusive)	Frequency	$S(d^2)$ (inclusive)	Frequency
0– 4	—	326–330	—
6– 14	—	316–324	—
16– 24	—	306–314	3
26– 34	2	296–304	5
36– 44	9	286–294	8
46– 54	20	276–284	18
56– 64	25	266–274	37
66– 74	32	256–264	36
76– 84	60	246–254	56
86– 94	63	236–244	44
96–104	90	226–234	83
106–114	104	216–224	102
116–124	110	206–214	92
126–134	117	196–204	120
136–144	96	186–194	139
146–154	125	176–184	134
156–164	129	166–174	141
Total	982	Total	1018

TABLE V

Distribution of 400 values of $S(d^2)$ for $n=20$

$S(d^2)$	Frequency
400–498	1
500–	0
600–	5
700–	10
800–	22
900–	22
1000–	40
1100–	44
1200–	51
1300–	56
1400–	42
1500–	35
1600–	33
1700–	17
1800–	8
1900–	8
2000–	4
2100–	2
Total	400

gives a chance of 0·9891 that a value of $S(d^2)$ will fall inside the range 38–292. The expected frequency outside this range in 2000 rankings is therefore 22, with a standard error of approximately $\sqrt{22}$. The observed frequency is 14. Similarly, for the 5 % level, the chance of a value falling inside the range 60–270 is 0·9474 and the expected frequency is thus 105; the observed frequency is 96.

Fig. 2 gives the histogram of the data of Table IV with the curve

$$y = k(1-x^2)^3$$

of equal range and equal area. So far as the eye can judge the correspondence is reasonably good.

Distribution of $S(d^2)$ in the rankings of 20

25. The distribution of $S(d^2)$ in the 400 rankings of 20 is given in a grouped form in Table V. The alternation of modes and antimodes has now disappeared, though it might emerge with finer grouping.

The mean value of $S(d^2)$ (about origin 1330), as grouped in Table V, is $-18·5$, the expected value being zero. The variance of the theoretical distribution is 93,100, so that the standard error of the mean of 400 sets is 15·2. Again the observed deviation is well within sampling limits.

The same is true of the variance, the deviation of the observed from the theoretical value being 714 with a standard error of 6193 approximately.

26. The experimental evidence, so far as it goes, confirms the theory. It would appear that for values of n equal to 10 or less the distribution of ρ cannot be taken to be normal to a satisfactory degree of approximation. The Type II B-curves proposed by Pitman are better, but they are possibly inadequate to represent frequencies in narrow ranges. They do, however, appear to be sufficient to determine significance points.

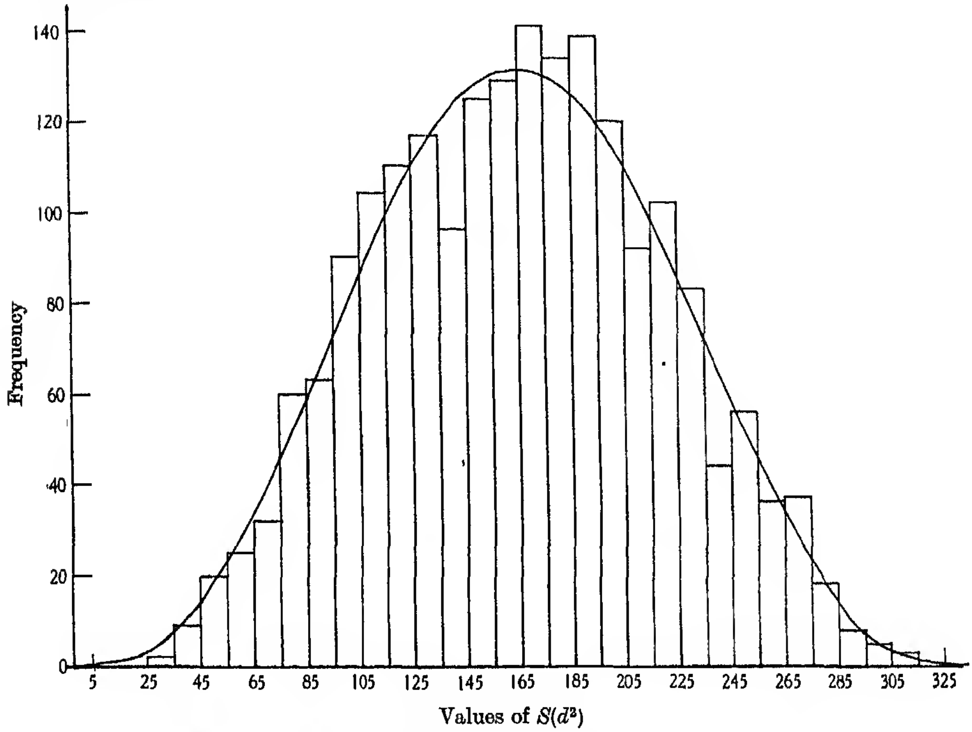


Fig. 2. Histogram of the data of Table IV, together with the curve $y=k(1-x^2)^3$ of equal area and corresponding range

PART III. RELATIONSHIP BETWEEN SPEARMAN'S COEFFICIENT AND ANOTHER COEFFICIENT OF RANK CORRELATION

27. One of us (Kendall, 1938) has suggested a measure of rank correlation whose sampling distribution can be obtained without much difficulty. For practical purposes the coefficient, denoted by τ, is most easily calculated as follows:

Let $X_1, X_2, \ldots, X_n$ be a permutation of the first n natural numbers. Suppose there are, to the right of X_1, k_1 numbers greater than X_1, to the right of X_2, k_2 numbers greater than X_2, and so on. If

$$\Sigma = 2S(k) - \tfrac{1}{2}n(n-1), \qquad \ldots\ldots(11)$$

the coefficient of rank correlation between the order X and the natural order $1, 2, \ldots, n$ is defined as

$$\tau = \frac{2\Sigma}{n(n-1)}. \qquad \ldots\ldots(12)$$

In the paper under reference it was shown that τ can vary from -1 to $+1$, has variance equal to $2(2n+5)/9n(n-1)$ in the universe in which all rankings appear equally frequently, and is normally distributed for large n. A method of obtaining the distribution for small n was given together with the actual distribution for values of n equal to 10 or less.

28. Different as ρ and τ might be expected to be from consideration of their methods of calculation, they were frequently found in practice to give numerical values which are remarkably close, even for low values of n. It therefore seemed worth while to investigate the relationship between them.

Each of the $n!$ permutations of the first n natural numbers will, in relation to the order $1, 2, \ldots, n$, give a pair of values of ρ and τ. The ideal would be to find the bivariate frequency table into which these values fall when arranged according to the values of ρ and τ (or, more conveniently, of $S(d^2)$ and Σ). Such a distribution must necessarily be extremely complicated when expressed in general terms, for even one of its border frequencies is the complex distribution of ρ (or $S(d^2)$).

It appears, however, to be possible to find a comparatively simple expression for the product-moment coefficient of correlation in such a table. This coefficient, denoted by $r_{\rho\tau}$ or $r_{S\Sigma}$ as the case may be, gives a reliable measure of the correspondence between ρ and τ inasmuch as the distribution of each is single-humped and tends to normality as n becomes larger.

29. By actually constructing the bivariate table for values of n from 2 to 6 inclusive we have found that, for such values,

$$r_{\rho\tau} = \frac{2(n+1)}{\sqrt{\{2n(2n+5)\}}}, \qquad \ldots\ldots(13)$$

with a corresponding value for the covariance of S and Σ,

$$\mu_{11} = -\tfrac{1}{18}n(n+1)^2(n-1). \qquad \ldots\ldots(14)$$

The actual correlation table for $n=6$ is given in Table VI. The close relation between the variates is immediately evident, and it is of some interest to note that the regression is not quite linear. Presumably, however, it approaches linearity as n becomes larger. Both variates tend to normality and though this in itself is insufficient to guarantee linearity of regression, the fact that $r_{\rho\tau}$ tends to unity makes it very probable that the joint distribution tends to the bivariate normal surface.

30. We have not succeeded in finding a rigid proof that equation (14) is true for all n. The following line of argument, however, appears to make it highly probable that (13) and (14) are of general application.

$\mu_{11}n!$, the product sum of Σ and $S(d^2)$ (the latter measured from its mean), is clearly an integer and is a function of n only; for when n is fixed it is completely

TABLE VI

Correlation table of $S(d^2)$ and Σ for $n=6$

Values of $S(d^2)$

Values of Σ	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	50	52	54	56	58	60	62	64	66	68	70	Total
15	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
13	.	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5
11	.	.	6	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	14
9	.	.	.	1	16	6	6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	29
7	.	.	.	.	.	6	8	20	11	.	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	49
5	.	.	.	.	.	.	.	4	8	21	10	22	4	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	71
3	.	.	.	.	.	.	.	.	1	.	9	4	20	32	8	8	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	90
1	.	.	.	.	.	.	.	.	.	.	.	2	.	2	12	22	30	13	16	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	101
−1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	16	13	30	22	12	2	.	2	.	.	.	.	.	.	.	.	.	.	.	101
−3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	8	8	32	20	4	9	.	1	.	.	.	.	.	.	.	.	90
−5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	4	22	10	21	8	4	.	.	.	.	.	.	.	71
−7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	.	11	20	8	6	.	.	.	.	.	49
−9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	6	16	1	.	.	.	29
−11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	6	.	.	14
−13	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	5
−15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
Totals	1	5	6	9	16	12	14	24	20	21	23	28	24	34	20	32	42	29	29	42	32	20	34	24	28	23	21	20	24	14	12	16	9	6	5	1	720

determined. The analogous quantities $\mu_{20}n!$ (the sum of squares of S) and $\mu_{02}n!$ (the sum of squares of Σ) are respectively

$$\frac{n^2(n+1)^2(n-1)}{36}n! \quad \text{and} \quad \frac{n(n-1)(2n+5)}{18}n!.$$

One suspects therefore that $\mu_{11}n!$ is equal to $f(n)n!$, where $f(n)$ is a polynomial in n; in other words, that

$$\mu_{11} = f(n).$$

If this is so, $f(n)$ cannot be of higher degree than four, for the product of μ_{20} and μ_{02} is of degree 8 and otherwise r would be greater than unity for some large n.

Hence if a polynomial of degree four or less can be found which takes the observed values of μ_{11} for five cases, that polynomial is equal to μ_{11}. Equation (14) satisfies the condition and thus is true in general.

In actual fact (14) is also satisfied in the degenerate case $n=1$, but (13) is not owing to the omission of two factors which cancel for $n>1$ but are zero for $n=1$.

31. If formula (13) is in fact true, the following are the values of r corresponding to some values of n:

n	$r_{\rho\tau}$
5	0·980
10	0·984
15	0·988
20	0·990

To verify the result for $n=10$ we found the coefficient of correlation between the values of ρ and τ for 1000 of the experimental permutations. This value was 0·980.

It would seem worthy of serious consideration, therefore, whether the coefficient ρ might not be replaced by τ, in the sampling distribution of which there is no uncertainty.

Summary

1. An expression is given for the sampling distribution of $S(d^2)$ in the universe in which all rankings appear equally frequently, where the Spearman coefficient of rank correlation is

$$\rho = 1 - \frac{6S(d^2)}{n^3-n}.$$

2. The distribution is given explicitly for values up to and including $n=8$.

3. It is suggested that for values of n less than 10 (and possibly higher) the distribution is inadequately represented by the normal curve but that a B-curve is sufficient to determine approximate significance points for values of n greater than 7.

4. Some experimental distributions for $n=10$ and $n=20$ are given and discussed. So far as they go these distributions support the theory.

5. A discussion is given of the relationship between ρ and a coefficient of rank correlation τ suggested elsewhere. The correlation between the two appears to be extremely high and in view of the fact that the sampling distribution of τ may be easily obtained it is suggested that τ may be of greater practical value than ρ.

REFERENCES

HOTELLING, HAROLD & PABST, MARGARET, R. (1936). "Rank Correlation and Tests of Significance Involving no Assumption of Normality." *Ann. Math. Statist.* **7**, 29.

KENDALL, M. G. & BABINGTON SMITH, BERNARD (1938). "Randomness and Random Sampling Numbers." *J.R. Statist. Soc.* **101**, 147

KENDALL, M. G. (1938). "A New Measure of Rank Correlation." *Biometrika*, **30**, 81.

PEARSON, KARL (1907). "On Further Methods of Determining Correlation." *Drap. Co. Mem. Biom. Ser.* IV, London, Dulau & Co.

PITMAN, E. J. G. (1937). "Significance Tests which may be applied to Samples from any Populations. II. The Correlation Coefficient Test." *J.R. Statist. Soc. Supp.* **4**, 225.

SPEARMAN, C. (1904). "The Proof and Measurement of Association between Two Things." *Amer. J. Psychol.* **15**, 88.

TIPPETT, L. H. C. (1927). Random Sampling Numbers, *Tracts for Computers, No.* 15, Cambridge University Press.

APPENDIX

The Randomness of the Experimental Samples

1. In testing the agreement between theory and the experimental data from Tippett's table we used some results obtained as follows:

Given a random series of n different objects, the average length of run required to reach one of $P\,(\leqslant n)$ stated objects is n/P.

For, if a start be made at any point in the series the chance that the first object is one of the P is P/n, say p. The chance that the first is not one of p but the second is so, is $(1-p)\,p$. The chance that the first $(r-1)$ are not members of P and that the rth is so, is $(1-p)^{r-1}\,p$.

The total chance of obtaining one of P is

$$p[1+(1-p)+(1-p)^2+\dots] = 1,$$

as it should.

The average length of run is

$$p[1+2(1-p)+3(1-p)^2+\dots]$$
$$=\frac{p}{[1-(1-p)]^2}=\frac{1}{p},\quad \text{the result stated.}$$

In other words, if there are at any stage P objects left to find to complete any given set, the average length of run required is n/P. Moreover the occurrence of each object is independent of that of the others. Hence the average length of run required to give $(n-1)$ of the n objects composing the series is

$$n\left[\frac{1}{n}+\frac{1}{n-1}+\frac{1}{n-2}+\dots+\frac{1}{3}+\frac{1}{2}\right]. \qquad \dots\dots(a)$$

In a similar way it will be seen that the variance (μ_2) of runs required to give one of P objects is given by

$$\mu_2+\frac{1}{p^2} = [1+2^2(1-p)+3^2(1-p)^2+\dots] = \frac{2}{p^2}-\frac{1}{p},$$

so that

$$\mu_2 = \frac{1}{p^2}-\frac{1}{p}.$$

Since the runs are independent the variance of the run required to give $(n-1)$ objects is

$$\mu_2 = n^2\left[\frac{1}{n^2}+\frac{1}{(n-1)^2}+\dots+\frac{1}{2^2}\right]-n\left[\frac{1}{n}+\frac{1}{n-1}+\dots+\frac{1}{2}\right]. \quad \dots\dots(b)$$

2. For the case $n=10$, formulae (*a*) and (*b*) give average run $=19{\cdot}29$, variance $=35{\cdot}69$. For 2000 sets the expected average value of the run is therefore 19·29 with a standard error of $\sqrt{\dfrac{35{\cdot}69}{2000}}$ or 0·134.

The observed value was 19·59, which exceeds the expected value by about 2·2 times the standard error.

This is rather too large for comfort. Possible sources of the difference are (*a*) the non-randomness of Tippett's numbers taken as a whole, (*b*) errors in writing down the permutations.

It appears that errors of type (*b*) would tend on the whole to exaggerate the length of run required since it is easier to overlook digits in the tables than to imagine non-existent digits. Such errors, however, unless they are systematically concerned with certain digits, which we regard as unlikely, will not affect the randomness of the permutations. Nevertheless, we thought it wise to eliminate this possible source of error in taking the sets of 20, and the method to this end has been described in the foregoing paper.

3. An internal test on the permutations of 10 themselves revealed no significant divergence from expectation. In one such test the numbers 1, 2, 3 were extracted from each permutation and their order noted. The results for the first 1920 permutations were:

Permutation	Observed frequency
123	302
132	296
213	339
231	327
312	323
321	333
Total	1920

The expected frequency in each class is 320, $\chi^2=4{\cdot}65$, $P=0{\cdot}46$ approx.

Moreover, as has been pointed out in the text, the resulting distribution of $S(d^2)$ conforms to expectation in its mean and variance.

4. Applying equations (*a*) and (*b*) when $n=20$ we find

average run $=$ 103·91 digits,

variance $=$ 746·04 (digits)^2.

The standard error of 400 sets is therefore 1·37.

The observed average run was 106·75, in excess of the theoretical run by 2·07 times the standard error.

This result confirms our suspicion that Tippett's numbers, taken as a whole, are not quite a suitable random set.

Nevertheless, the resulting distribution of $S(d^2)$ conforms to expectation in mean and variance.

5. To sum up, we are inclined to suspect that Tippett's numbers may give results not in accordance with expectation when the whole table is used. The difference, however, is not greatly beyond permissible sampling limits. Moreover the non-randomness of Tippett's table, even if it exists, need not necessarily affect the randomness of the permutations obtained from it or of the calculated value of $S(d^2)$, and internal tests suggest that, in fact, it has not done so. We feel, therefore, that the sample of values of $S(d^2)$ may be regarded as random with some considerable confidence.

THE APPLICATION OF THE MOMENT FUNCTION IN THE STUDY OF DISTRIBUTION LAWS IN STATISTICS*

By U. S. NAIR

Department of Mathematics, Travancore University, South India

CONTENTS

1. Introduction

In the study of distribution laws in statistics the two most important methods are those depending upon

(a) The characteristic function;

(b) Transformation of variables.

Sometimes both these methods lead to certain definite integrals which are not capable of being expressed in terms of simple functions. In such cases it is a common practice to approximate the distribution by means of the very well-known system of curves due to Karl Pearson.

In the present paper a method of deriving distribution laws from a slightly different point of view is developed. Certain theorems regarding this method are proved in § 2, and the remaining sections are devoted to the application of these theorems to derive the distribution of several criteria that arise in the Theory of Sampling.

The illustrations taken have been partly studied by S. S. Wilks (1932), who expresses the distribution laws as multiple integrals which may readily be evaluated in certain simple cases. The present method however gives the result as a single integral whose properties, from the mathematical point of view, may be studied by means of a differential equation that it satisfies.

* The present paper is a modification of one of the papers submitted by the author for the Ph.D. Degree in Statistics of the University of London (1937).

The distribution laws derived in the paper may appear as if they are pure mathematical functions which cannot with advantage be handled by the practical statistician. No doubt the distribution laws take a very complicated form, but the author has taken the formula of § 3 and shown how this may conveniently be used to calculate the levels of significance of the L_1 criterion. By this method he has been able to check the substantial accuracy of the 5 % and 1 % significance levels of the L_1 criterion obtained and tabled by P. P. N. Nayer (1936) by an approximate method. A further paper setting out these results and discussing their bearing on the accuracy of an analogous test suggested by M. S. Bartlett (1937*a*, *b*) will be published shortly. The other distribution laws may be utilized on similar lines to yield useful information, which otherwise would be lacking.

2. Certain theorems regarding distribution functions

(1) It is proposed to develop in this section a few theorems yielding distribution functions. The method is based on the theory of Fourier's transform, and the formula developed is due to Mellin.

To avoid repetition we shall adopt the following notation:

(*a*) $x_1, x_2, \ldots, x_n$ are n variates continuous in the interval $a_i \leqslant x_i \leqslant b_i$ ($i = 1, 2, \ldots, n$); a_i and b_i may be finite or infinite.

(*b*) $p(x_1, x_2, \ldots, x_n)$ is the probability law of the x's so that

$$\int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n = 1. \qquad \ldots\ldots(1)$$

(*c*) $\theta_1, \theta_2, \ldots, \theta_n$ are non-negative functions of the x's. Any one of these will be denoted by θ.

(2) THEOREM 1. *If*

$$\phi(t) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} \theta^t p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n, \qquad \ldots\ldots(2)$$

then

$$p(\theta) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \theta^{-t-1} \phi(t)\, dt, \qquad \ldots\ldots(3)$$

provided the integrals in (2) *and* (3) *exist.**

Proof. To prove this theorem we note that if $u = \log\theta$,

$$p(\theta) = p(u)\frac{du}{d\theta} = \theta^{-1} p(u). \qquad \ldots\ldots(4)$$

We have

$$\phi(it) = \int \cdots \int \theta^{it} p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n$$

$$= \int_{a_1}^{a_1} \cdots \int_{a_n}^{b_n} e^{itu} p(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n. \qquad \ldots\ldots(5)$$

* It will be seen that for convergence of the integrals in (2) and (3), it is enough that $\phi(t)$ belongs to L_2. The inversion formula (3) holds true even if $\phi(t)$ belongs to L_1.

Hence, using Fourier's theorem,

$$p(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itu}\,\phi(it)\,dt. \qquad \ldots\ldots(6)$$

Changing u to θ and it to t, we get

$$p(\theta) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta^{-t-1}\,\phi(t)\,dt. \qquad \ldots\ldots(7)$$

THEOREM 2. *If*

$$\phi(t_1, t_2) = \int_{a_1}^{b_1}\cdots\int_{a_2}^{b_2} \theta_1^{t_1}\theta_2^{t_2} p(x_1, x_2, \ldots, x_n)\,dx_1 dx_2 \ldots dx_n, \qquad \ldots\ldots(8)$$

then

$$p(\theta_1, \theta_2) = -\frac{1}{4\pi^2}\int_{-i\infty}^{+i\infty}\int_{-i\infty}^{i\infty} \theta_1^{-t_1-1}\theta_2^{-t_2-1}\,\phi(t_1, t_2)\,dt_1 dt_2, \qquad \ldots\ldots(9)$$

provided the integrals in (8) *and* (9) *exist.*

This theorem may be proved by the same method adopted for theorem 1, but we give below a slightly different proof.

Proof. Let $p(\theta_2 \mid \theta_1)$ denote the probability law of θ_2 given θ_1, so that

$$p(\theta_1, \theta_2) = p(\theta_1)\,p(\theta_2 \mid \theta_1). \qquad \ldots\ldots(10)$$

Now

$$\phi(t_1, t_2) = \int_{\alpha_1}^{\beta_1}\int_{\alpha_2}^{\beta_2} \theta_1^{t_1}\theta_2^{t_2} p(\theta_1, \theta_2)\,d\theta_1 d\theta_2$$

$$= \int_{\alpha_1}^{\beta_1} \theta_1^{t_1}\left[\int_{\gamma}^{\delta} \theta_2^{t_2} p(\theta_2 \mid \theta_1)\,d\theta_2\right] p(\theta_1)\,d\theta_1 \qquad \ldots\ldots(11)$$

$$= \int_{\alpha_1}^{\beta_1} \theta_1^{t_1}\,g(t_2)\,p(\theta_1)\,d\theta_1, \qquad \ldots\ldots(12)$$

where

$$g(t_2) = \int_{\gamma}^{\delta} \theta_2^{t_2}\,p(\theta_2 \mid \theta_1)\,d\theta_1. \qquad \ldots\ldots(13)$$

In the integrals in (11) α_1 and β_1, α_2 and β_2 are the limits of θ_1 and θ_2, and these may be finite or infinite. γ and δ in (13) may depend on θ_1.

Applying the inversion formula of theorem 1 to (12) and (13), we obtain

$$p(\theta_1)\,g(t_2) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta_1^{-t_1-1}\,\phi(t_1, t_2)\,dt_1, \qquad \ldots\ldots(14)$$

$$p(\theta_2 \mid \theta_1) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta_2^{-t_2-1}\,g(t_2)\,dt_2. \qquad \ldots\ldots(15)$$

From (14) and (15) we get

$$p(\theta_2 \mid \theta_1)\,p(\theta_1) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta_2^{-t_2-1}\left[\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta_1^{-t_1-1}\,\phi(t_1, t_2)\,dt_1\right] dt_2$$

$$= \frac{-1}{4\pi^2}\int_{-i\infty}^{i\infty}\int_{-i\infty}^{i\infty} \theta_1^{-t_1-1}\theta_2^{-t_2-1}\,\phi(t_1, t_2)\,dt_1 dt_2. \qquad \ldots\ldots(16)$$

Hence the theorem follows from (10).

THEOREM 3. *If*

$$\psi(t_1, t_2) = \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} e^{it_1\theta_1}\theta_2^{t_2}\, p(x_1, x_2, x_3, \dots, x_n)\, dx_1 dx_2 \dots dx_n, \qquad \dots\dots(17)$$

$$p(\theta_1, \theta_2) = \frac{1}{4\pi^2 i}\int_{-\infty}^{\infty} e^{-it_1\theta_1}\, dt_1 \left[\int_{-i\infty}^{i\infty} \theta_2^{-t_2-1}\psi(t_1, t_2)\, dt_2\right], \qquad \dots\dots(18)$$

provided the integrals (17) *and* (18) *exist.*

The proof of this theorem is obvious. It is also clear that the theorems 2 and 3 may be extended to give the simultaneous distribution functions of any number of variates.

It will be observed that $\phi(t)$ in theorem 1 gives the mathematical expectation of the tth power of θ, and hence for positive integral values of t, $\phi(t)$ is the tth moment of θ. But the integral (2) defines $\phi(t)$ for all values of t for which the integral exists. Hence it is proposed to call $\phi(t)$ the moment function of θ.

(3) The integral (3) may conveniently be evaluated by the method of contour integration. But in certain cases it is found that a very easy method is afforded by considering it as the solution of a differential equation. This method is given below.*

THEOREM 4. *Suppose $\phi(t)$ in theorem 1 satisfies the following conditions:*

(*a*) *The singularities of $\phi(t)$ are all on the negative axis of t.*

(*b*) *There is a positive number a such that*

$$\phi(t+a) = \frac{A(t)}{B(t)}\phi(t), \qquad \dots\dots(19)$$

$A(t)$ and $B(t)$ being polynomials in t.

Under these conditions $p(\theta)$ defined by (3) *is a solution of the differential equation*

$$A\left(-\theta\frac{d}{d\theta}-1\right)z - \theta^a B\left(-\theta\frac{d}{d\theta}-a-1\right)z = 0. \qquad \dots\dots(20)$$

Proof. Since the singularities of $\phi(t)$ are on the negative axis of t, we may move the path of integration in (3) to $(a-i\infty, a+i\infty)$ without changing the value of the integral. Thus the equation (3) reduces to

$$p(\theta) = \frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} \theta^{-t-1}\phi(t)\, dt. \qquad \dots\dots(21)$$

Replacing t by $t+a$ and using (19), we have

$$p(\theta) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta^{-t-a-1}\phi(t+a)\, dt$$

$$= \theta^{-a}\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \theta^{-t-1}\frac{A(t)}{B(t)}\phi(t)\, dt. \qquad \dots\dots(22)$$

* I am grateful to Dr G. Rasch for pointing out this method in the study of distribution functions.

Now we use the following symbolic equations:*

$$f\left(\theta\frac{d}{d\theta}\right)\theta^n = f(n)\,\theta^n, \qquad \text{......(23)}$$

$$f\left(\theta\frac{d}{d\theta}\right)\theta^n\gamma = \theta^n f\left(\theta\frac{d}{d\theta}+n\right)\gamma. \qquad \text{......(24)}$$

From (23),
$$\theta^{-t-1}A(t) = A\left(-\theta\frac{d}{d\theta}-1\right)\theta^{-t-1}. \qquad \text{......(25)}$$

Substituting the result of (25) in (22) and assuming that the symbolic operator can be taken out of the integral sign, we have

$$\theta^a p(\theta) = A\left(-\theta\frac{d}{d\theta}-1\right)\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\theta^{-t-1}\frac{1}{B(t)}\phi(t)\,dt. \qquad \text{......(26)}$$

Performing the operation $B\left(-\theta\frac{d}{d\theta}-1\right)$ on both sides of (26) and after simplification with the help of (24), we get

$$\theta^a B\left(-\theta\frac{d}{d\theta}-a-1\right)p(\theta) = A\left(-\theta\frac{d}{d\theta}-1\right)p(\theta), \qquad \text{......(27)}$$

which proves that $p(\theta)$ satisfies (20).

The solution of the differential equation (20) will have a certain number of constants equal to the order of the equation. By a proper choice of these constants the solution may be made identical to $p(\theta)$. This may be done by equating the residue of the integrand in (3) at any of its poles to the corresponding terms in the solution of (20).

The method of the differential equation may generally be employed when $\phi(t)$ is of the form

$$\phi(t) = \prod_{i=1}^{n}\frac{\Gamma(a_i+t)\,\Gamma(b_i)}{\Gamma(b_i+t)\,\Gamma(a_i)}, \qquad \text{......(28)}$$

where a_i is greater than zero for $i = 1, 2, \ldots, n$; for, in this case, the singularities of $\phi(t)$ are the same as those of $\Gamma(a_i+t)$ and these are at the points $t = -j-a_i$ where j is zero or any positive integer. Also

$$\phi(t+1) = \phi(t)\prod_{i=1}^{n}\frac{(a_i+t)}{(b_i+t)}.$$

Thus we have here a method of studying a certain type of integral equation that has been called by S. S. Wilks (1932, p. 474) a type B integral equation.

* See, for example, A. R. Forsyth (1921), theorems 1 and 2, p. 61.

3. THE SAMPLING DISTRIBUTION OF THE NEYMAN-PEARSON L_1 CRITERION IN THE CASE OF k SAMPLES OF EQUAL SIZE

(1) Let $s_1, s_2, \ldots, s_k$* be the sample standard deviations from k normal populations with the same standard deviation. Then L_1 is defined by Neyman & Pearson (1931) as

$$L_1 = \frac{(s_1^2 s_2^2 \ldots s_k^2)^{1/k}}{\left(\dfrac{s_1^2+s_2^2+\ldots+s_k^2}{k}\right)}. \qquad \ldots\ldots(29)$$

If $\phi(t)$ is the moment function of L_1,

$$\phi(t) = \int_0^\infty \ldots \int_0^\infty L_1^t p(s_1^2, s_2^2, \ldots, s_k^2)\, ds_1^2 \ldots ds_k^2$$

$$= \frac{\Gamma\left(k\dfrac{n-1}{2}\right)}{\Gamma^k\left(\dfrac{n-1}{2}\right)} \frac{\Gamma^k\left(\dfrac{t}{k}+\dfrac{n-1}{2}\right)}{\Gamma\left(t+k\dfrac{n-1}{2}\right)} k^t. \qquad \ldots\ldots(30)$$

Now, we apply the inversion formula of theorem 1 and get

$$p(L_1) = \frac{\Gamma\left(k\dfrac{n-1}{2}\right)}{\Gamma^k\left(\dfrac{n-1}{2}\right)} \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} L_1^{-t-1} \frac{\Gamma^k\left(\dfrac{t}{k}+\dfrac{n-1}{2}\right)}{\Gamma\left(t+k\dfrac{n-1}{2}\right)} k^t dt. \qquad \ldots\ldots(31)$$

To evaluate this integral, replace $\dfrac{t}{k}+\dfrac{n-1}{2}$ by $\dfrac{t}{k}$. Then

$$p(L_1) = \frac{\Gamma\left(k\dfrac{n-1}{2}\right)}{\Gamma^k\left(\dfrac{n-1}{2}\right)} \frac{L_1^{\frac{1}{2}k(n-1)-1}}{k^{\frac{1}{2}k(n-1)}} \frac{1}{2\pi i}\int_{\frac{1}{2}(n-1)-i\infty}^{\frac{1}{2}(n-1)+i\infty} L_1^{-t} \frac{\Gamma^k(t/k)}{\Gamma(t)} k^t dt. \ldots\ldots(32)$$

Clearly the poles of the integrand are the same as those of $\Gamma^k(t/k)$ and, except for the pole at the origin, these lie on the negative axis of t. The path of integration may therefore be changed to $c-i\infty$, $c+i\infty$, where c is any positive number. Hence we write

$$p(L_1) = F_1(L_1)\, F_2(L_1), \qquad \ldots\ldots(33)$$

where

$$F_1(L_1) = \frac{\Gamma\left(k\dfrac{n-1}{2}\right)}{\Gamma^k\left(\dfrac{n-1}{2}\right)} \frac{L_1^{\frac{1}{2}k(n-1)-1}}{k^{\frac{1}{2}k(n-1)}}, \qquad \ldots\ldots(34)$$

$$F_2(L_1) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left(\frac{L_1}{k}\right)^{-t} \frac{\Gamma^k(t/k)}{\Gamma(t)} dt. \qquad \ldots\ldots(35)$$

* If $x_1, x_2, \ldots, x_n$ are the sample values, s^2 is the mean value of $(x_i-\bar{x})^2$, $\bar{x}$ being the arithmetic mean of the x's.

Thus $p(L_1)$ splits into the product of two factors, one of which is independent of the size of the sample.

Putting L_1/k equal to x and noting that, if $\chi(t) = \dfrac{\Gamma^k(t/k)}{\Gamma(t)}$

$$\chi(t+k) = \frac{t^{k-1}}{k^k(t+1)(t+2)\dots(t+k-1)}\chi(t), \qquad \text{......(36)}$$

we get from theorem 4 the following differential equation satisfied by $F_2(L_1)$:

$$(xk)^k\left(x\frac{d}{dx}+1\right)\left(x\frac{d}{dx}+2\right)\dots\left(x\frac{d}{dx}+k-1\right)z = \left(x\frac{d}{dx}\right)^{k-1}z. \qquad \text{......(37)}$$

(2) To solve equation (37), assume

$$z = \sum_{i=0}^{\infty} a_i x^{\rho+i}. \qquad \text{......(38)}$$

Substituting in (37)

$$k^k\sum_{i=0}^{\infty} a_i(\rho+i+1)(\rho+i+2)\dots(\rho+i+k-1)\,x^{\rho+i+k}$$
$$= \sum_{i=0}^{\infty} a_i(\rho+i)^{k-1}x^{\rho+i}. \qquad \text{......(39)}$$

Equating to zero the lowest power of x, the indicial equation is

$$\rho^{k-1} = 0, \qquad \text{......(40)}$$

which gives $\rho = 0$ as a $(k-1)$ multiple root. Further, the coefficients satisfy the recurrence formula

$$a_{i+k} = a_i k^k \frac{\Gamma(\rho+i+k+1)}{\Gamma(\rho+i+1)(\rho+i+k)^k}, \qquad \text{......(41)}$$

with $a_i = 0$ for $i = 1, 2, \dots, (k-1)$.

Thus the series (38) contains only terms of the type x^{ki} and we may write

$$z = \sum_{i=0}^{\infty} A_i(\rho)\,x^{ki+\rho}, \qquad \text{......(42)}$$

where
$$A_i(\rho) = k^{ki}\frac{\Gamma(\rho+ik+1)}{\Gamma(\rho+1)[(\rho+k)(\rho+2k)\dots(\rho+ik)]^k}A_0. \qquad \text{......(43)}$$

To get the complete solution, since the indicial equation has $\rho = 0$ as a $(k-1)$ multiple root, we use the method of Frobenius and obtain

$$z = \sum_{h=0}^{k-2} C_h\left(\frac{\partial}{\partial\rho}\right)^h_{\rho=0}\Sigma A_i(\rho)\,x^{ki+\rho}, \qquad \text{......(44)}$$

where C_h $(h = 0, 1, \dots, k-2)$ are constants.

It may easily be proved that the series (42) is uniformly convergent in the interval $0 \leqslant \rho$ and $0 \leqslant xk \leqslant 1$. Hence differentiation term by term of this series s valid.

To evaluate the hth derivative of $A_i(\rho)\, x^{ik+\rho}$ we write

$$A_i(\rho)\, x^{ik+\rho} = (kx)^{ik}\, e^{\rho\log x+\log\Gamma(\rho+ik+1)-\log\Gamma(\rho+1)-k\log(\rho+k)(\rho+2k)\ldots(\rho+ik)}. \qquad \ldots\ldots(45)$$

Using the formula

$$\log\Gamma(x+a) = \log\Gamma(a)+x\psi(a)+\frac{x^2}{2!}\psi_1(a)+\frac{x^3}{3!}\psi_2(a)+\ldots, \qquad \ldots\ldots(46)$$

where

$$\left.\begin{aligned} \psi(a) &= \frac{d}{dx}\log\Gamma(x)|_{x=a}, \\ \psi_j(a) &= \left(\frac{d}{dx}\right)^j \psi(x)|_{x=a}, \end{aligned}\right\} \qquad \ldots\ldots(47)$$

the exponent in (45) may be expanded and we get

$$A_i(\rho)\, x^{ik+\rho} = \frac{(ik)!}{(i!)^k} x^{ik} \exp\left[b_1(i)\,\rho+b_2(i)\frac{\rho^2}{2!}+b_3(i)\frac{\rho^3}{3!}+\ldots\right], \qquad \ldots\ldots(48)$$

where

$$\left.\begin{aligned} b_1(i) &= \psi(ik+1)-\psi(1)-\left(1+\frac{1}{2}+\ldots+\frac{1}{i}\right)+\log x, \\ b_j(i) &= \psi_{j-1}(ik+1)-\psi_{j-1}(1)+(-1)^j\left(1+\frac{1}{2^j}+\ldots+\frac{1}{i^j}\right)\frac{(j-1)!}{k^{j-1}}. \end{aligned}\right\} \ldots(49)$$

Hence

$$\left(\frac{\partial}{\partial\rho}\right)^h_{\rho=0} A_i(\rho)\, x^{ik+\rho} = \frac{(ik)!}{(i!)^k} D_h(i)\, x^{ik}, \qquad \ldots\ldots(50)^*$$

where

$$D_h(i) = \begin{vmatrix} b_1(i) & -1 & 0 & 0 & 0 & \ldots & 0 \\ b_2(i) & b_1(i) & -1 & 0 & 0 & \ldots & 0 \\ b_3(i) & 2b_2(i) & b_1(i) & -1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ b_h(i) & \binom{h-1}{1} b_{h-1}(i) & \binom{h-1}{2} b_{h-2}(i) & \ldots & \ldots & \ldots & b_1(i) \end{vmatrix}, \qquad \ldots\ldots(51)$$

$\binom{h}{r}$ denoting the binomial coefficient ${}^hC_r$. Hence

$$z = \sum_{h=0}^{k-2} C_h \sum_{i=0}^{\infty} \frac{(ik)!}{(i!)^k} D_h(i)\, x^{ik}. \qquad \ldots\ldots(52)$$

To evaluate the constants C_h in (52) so that z becomes identical with $F_2(L_1)$, we find the residue of $x^{-t}\dfrac{\Gamma^k(t/k)}{\Gamma(t)}$ at $t=0$ and equate this to the corresponding terms in (52). Now

$$x^{-t}\frac{\Gamma^k(t/k)}{\Gamma(t)} = x^{-t}\frac{\Gamma^k(t/k+1)}{\Gamma(t+1)}\,\frac{k^k}{t^{k-1}}. \qquad \ldots\ldots(53)$$

* See Appendix.

Hence at $t = 0$ the pole is of order $k-1$ and the residue is the coefficient of t^{k-2} in the expansion of

$$x^{-t}\,k^k \frac{\Gamma^k(t/k+1)}{\Gamma(t+1)}$$

which is equal to

$$\frac{k^k}{(k-2)!}\left(\frac{d}{dt}\right)^{k-2}_{t=0} x^{-t}\frac{\Gamma^k(t/k+1)}{\Gamma(t+1)} = \frac{k^k}{(k-2)!}\left(\frac{d}{dt}\right)^{k-2}_{t=0} \exp\left[-t\log x + d_2\frac{t^2}{2!} + d_3\frac{t^3}{3!} + \dots\right]$$

$$= \frac{k^k}{(k-2)!}\begin{vmatrix} -\log x & -1 & 0 & 0 & \dots & 0 \\ d_2 & -\log x & -1 & 0 & \dots & 0 \\ d_3 & 2d_2 & -\log x & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{k-2} & \binom{k-3}{1} d_{k-3} & \binom{k-3}{2} d_{k-4} & \dots & \dots & -\log x \end{vmatrix}, \qquad \dots\dots(54)$$

where $$d_j = \frac{1-k^{j-1}}{k^{j-1}}\,\psi_{j-1}(1) \quad (j = 2, 3, \dots).$$

The corresponding terms in (52) are clearly

$$\sum_{h=0}^{k-2} C_h D_h(0). \qquad \dots\dots(55)$$

Since $$b_1(0) = \log x \quad \text{and} \quad b_2(0) = b_3(0) = \dots = 0,$$
(55) reduces to

$$\sum_{h=0}^{k-2} C_h D_h(0) = C_0 + C_1 \log x + C_2(\log x)^2 + \dots + C_{k-2}(\log x)^{k-2}. \qquad \dots\dots(56)$$

Equating (56) to (54) and putting $\log x = -\theta$, the constants are given by the equation

$$C_0 - C_1\theta + C_2\theta^2 \dots + (-1)^{k-2} C_{k-2}\theta^{k-2}$$

$$= \frac{k^k}{(k-2)!}\begin{vmatrix} \theta & -1 & 0 & 0 & 0 & \dots & 0 \\ d_2 & \theta & -1 & 0 & 0 & \dots & 0 \\ d_3 & 2d_2 & \theta & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d_{k-2} & \binom{k-3}{1} d_{k-3} & \binom{k-3}{2} d_{k-4} & \dots & \dots & \dots & \theta \end{vmatrix}. \qquad \dots\dots(57)$$

Hence $$C_0 + C_1 D_1(i) + C_2 D_2(i) + C_3 D_3(i) \dots + C_{k-2} D_{k-2}(i)$$

$$= \frac{k^k}{(k-2)!}\Delta(i, k), \qquad \dots\dots(58)$$

where

$$\Delta(i,k) = \begin{vmatrix} d_1(i,k) & -1 & 0 & 0 & \ldots & 0 \\ d_2(i,k) & d_1(i,k) & -1 & 0 & \ldots & 0 \\ d_3(i,k) & 2d_2(i,k) & d_1(i,k) & -1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ d_{k-2}(i,k) & \binom{k-3}{1} d_{k-3}(i,k) & \binom{k-3}{2} d_{k-4}(i,k) & \ldots & \ldots & d_1(i,k) \end{vmatrix}, \qquad \ldots\ldots(59)^*$$

the elements of the determinant being defined by the following equations:

$$\left.\begin{aligned} d_1(i,k) &= b_1(i), \\ d_j(i,k) &= b_j(i) + (-1)^j \left(1 - \frac{1}{k^j}\right) \psi_j(1). \end{aligned}\right\} \qquad \ldots\ldots(60)$$

Thus

$$F_2(L_1) = \frac{k^k}{(k-2)!} \sum_{i=0}^{\infty} \frac{(ik)!}{(i!)^k k^{ki}} \Delta(i,k) L_1^{ik}. \qquad \ldots\ldots(61)$$

Finally, from equations (34), (35) and (61),

$$p(L_1) = \frac{\Gamma\left(k\frac{n-1}{2}\right)}{\Gamma^k\left(\frac{n-1}{2}\right)} \frac{L_1^{\frac{1}{2}k(n-1)-1}}{(k-2)!\, k^{\frac{1}{2}k(n-3)}} \sum_{i=0}^{\infty} \frac{(ik)!}{(i!)^k k^{ki}} \Delta(i,k) L_1^{ik}. \qquad \ldots\ldots(62)\dagger$$

When $k = 2$, the differential equation reduces to

$$(1-4x^2)\frac{dz}{dx} - 4xz = 0.$$

This is easily integrated and we get

$$z = C(1-4x^2)^{-\frac{1}{2}}. \qquad \ldots\ldots(63)$$

The constant C is given by

$$\begin{aligned} C &= \text{residue of } x^{-t}\frac{\Gamma^2(\frac{1}{2}t)}{\Gamma(t)} \text{ at } t = 0 \\ &= 4. \end{aligned}$$

Hence in this special case

$$p(L_1) = \frac{\Gamma(n-1)}{\Gamma^2\left(\frac{n-1}{2}\right)} \frac{L_1^{n-2}}{2^{n-3}} (1-L_1^2)^{-\frac{1}{2}}, \qquad \ldots\ldots(64)$$

a result which has been proved by P. P. N. Nayer (1936, p. 43).

4. The sampling distribution of the L_1 criterion appropriate to k samples of equal size from bi-variate normal populations

(1) In this section we shall consider the sampling distribution of the likelihood criterion appropriate for k samples from bi-variate normal populations as developed by S. S. Wilks (1932, pp. 489–90). We shall denote this criterion by λ_1, instead of using Wilks' notation of $\lambda_{(H'n)}$, n being the number of variates. It will

* See Appendix.

† The application of this series to numerical computation will be dealt with in a separate paper.

be observed that in the general case this gives a single criterion to test the hypothesis that the variances and the co-variances in a set of k multivariate normal populations are equal.

Consider k p-variate normal populations.* Let

(1) $x_1, x_2, \ldots, x_p$ be the p variates.

(2) $n_1, n_2, \ldots, n_k$ the size of the k samples $(n_1+n_2+\ldots+n_k = N)$.

(3) x_{iab} the value of the ith variate for the ath individual in the bth sample.

(4) $\bar{x}_{ib}$ the mean of the ith variate in the bth sample.

(5) $s_{ijb} = \dfrac{1}{n_b}\sum\limits_{a=1}^{n_b}(x_{iab}-\bar{x}_{ib})(x_{jab}-\bar{x}_{jb})$.

(6) $c_{ij} = \dfrac{1}{N}\sum\limits_{b=1}^{k} n_b s_{ijb}$.

(7) v_b the generalized variance for the sample b, i.e. the determinant $|s_{ijb}|$.

(8) v the generalized variance for the sample obtained by pooling together all the k samples, i.e. $v = |c_{ij}|$.

With this notation
$$\lambda_1 = \prod_{b=1}^{k}\left(\frac{v_b}{v}\right)^{\frac{1}{2}n_b}. \qquad \ldots\ldots(65)$$

(2) The tth moment of λ_1 is given by Wilks (1932, p. 490) as

$$\phi(t) = \prod_{b=1}^{k}\left[\left(\frac{N}{n_b}\right)^{\frac{1}{2}tn_b p}\prod_{i=1}^{p}\frac{\Gamma\left(\dfrac{n_b(1+t)-i}{2}\right)}{\Gamma\left(\dfrac{n_b-i}{2}\right)}\right]\prod_{i=1}^{p}\frac{\Gamma\left(\dfrac{N-k+1-i}{2}\right)}{\Gamma\left(\dfrac{N(1+t)-k+1-i}{2}\right)}. \qquad \ldots\ldots(66)$$

When the n's are equal, (66) reduces to

$$\phi(t) = \prod_{i=1}^{p}\frac{\Gamma\left\{\dfrac{k(n-1)-i+1}{2}\right\}}{\Gamma^k\left(\dfrac{n-i}{2}\right)}\,k^{\frac{1}{2}(npkt)}\prod_{i=1}^{p}\frac{\Gamma^k\left\{\dfrac{n-i}{2}+\dfrac{nt}{2}\right\}}{\Gamma\left\{\dfrac{k(n-1)-i+1}{2}+\dfrac{nkt}{2}\right\}}. \qquad \ldots\ldots(67)$$

Applying the inversion formula of theorem 1 to (67), we get†

$$p(\lambda_1) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\lambda_1^{-t-1}\phi(t)\,dt$$
$$= \prod_{j=1}^{p}\frac{\Gamma\left\{\dfrac{k(n-1)-j+1}{2}\right\}}{\Gamma^k\left(\dfrac{n-j}{2}\right)}\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\lambda_1^{-t-1}\,k^{\frac{1}{2}(npkt)}\prod_{j=1}^{p}\frac{\Gamma^k\left(\dfrac{n-j}{2}+\dfrac{nt}{2}\right)}{\Gamma\left(\dfrac{k(n-1)-j+1}{2}+\dfrac{nkt}{2}\right)}\,dt. \qquad \ldots\ldots(68)$$

* I have followed Wilks' notation in using the letter p for the number of variates. This must not be confused with the p used to denote a probability function.

† By applying Stirling's formula for $\Gamma(x)$, we may prove that $\phi(t) \sim O(t^{-\frac{1}{4}(k-1)p(p+1)})$ and hence the integral (68) exists.

Putting $t = \dfrac{\theta}{nk}$, (68) becomes

$$p(\lambda_1) = \prod_{j=1}^{p} \frac{\Gamma\left\{\dfrac{k(n-1)-j+1}{2}\right\}}{\Gamma^k\left(\dfrac{n-j}{2}\right)} \frac{1}{nk} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \lambda_1^{-\theta/nk-1} k^{\frac{1}{2}p\theta} \prod_{j=1}^{p} \frac{\Gamma^k\left(\dfrac{n-j}{2}+\dfrac{\theta}{2k}\right)}{\Gamma\left(\dfrac{k(n-1)-j+1}{2}+\dfrac{\theta}{2}\right)} d\theta. \qquad \text{......(69)}$$

Changing λ_1 to $L_1 = \lambda_1^{1/nk}$ and using the relation

$$p(L_1) = p(\lambda_1) L_1^{nk-1} nk,$$

we get

$$p(L_1) = \prod_{j=1}^{p} \frac{\Gamma\left\{\dfrac{k(n-1)-j+1}{2}\right\}}{\Gamma^k\left(\dfrac{n-j}{2}\right)} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} L_1^{-\theta-1} k^{\frac{1}{2}p\theta} \prod_{j=1}^{p} \frac{\Gamma^k\left(\dfrac{n-j}{2}+\dfrac{\theta}{2k}\right)}{\Gamma\left(\dfrac{k(n-1)-j+1}{2}-\dfrac{\theta}{2}\right)} d\theta. \qquad \text{......(70)}$$

Now putting $\dfrac{n-p}{2}+\dfrac{\theta}{2k} = \dfrac{t}{2k}$ we get

$$p(L_1) = F_1(L_1) F_2(L_1), \qquad \text{......(71)}$$

where

$$F_1(L_1) = \prod_{j=1}^{p} \frac{\Gamma\left\{\dfrac{k(n-1)-j+1}{2}\right\}}{\Gamma^k\left(\dfrac{n-j}{2}\right)} \frac{L_1^{k(n-p)-1}}{k^{kp\frac{1}{2}(n-p)}}, \qquad \text{......(72)}$$

$$F_2(L_1) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} L_1^{-t} k^{\frac{1}{2}(pt)} \prod_{j=1}^{p} \frac{\Gamma^k\left(\dfrac{p-j}{2}+\dfrac{t}{2k}\right)}{\Gamma\left(\dfrac{k(p-1)-j+1}{2}+\dfrac{t}{2}\right)} dt. \qquad \text{......(73)}$$

Thus $p(L_1)$ splits into the product of two factors one of which is independent of the size of the sample. This result may be compared with that of equation (33).

To evaluate the integral for $F_2(L_1)$, we might apply the method of theorem 4; this however is not particularly easy. We shall therefore apply the direct method of contour integration. It is evident that the poles of the integrand in (73) are given by

$$\frac{p-j}{2}+\frac{t}{2k} = -s \quad \text{for } s = 0, 1, 2, \ldots, \infty;\ j = 1, 2, 3, \ldots, p. \qquad \text{......(74)}$$

The occurrence of the double suffix s and j makes the actual expression for the residue highly involved. So we shall limit the discussion to the case $p = 2$ and show how the expression for the integral may be reduced to a manageable form.

When $p = 2$ we get, after simplification, with the help of the formula

$$2^{2x-1} \Gamma(x) \Gamma(x+\tfrac{1}{2}) = \sqrt{\pi}\, \Gamma(2x), \qquad \text{......(75)}$$

$$p(L_1) = F_1(L_1) F_2(L_1), \qquad \text{......(76)}$$

where
$$F_1(L_1) = \frac{\Gamma(k\overline{n-1}-1)}{\Gamma^k(n-2)}\frac{L_1^{k(n-2)-1}}{k^{k(n-2)}}, \qquad \ldots\ldots(77)$$

$$F_2(L_1) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\left(\frac{L_1}{k}\right)^{-t}\frac{\Gamma^k(t/k)}{\Gamma(t+k-1)}\,dt. \qquad \ldots\ldots(78)$$

In $F_2(L_1)$ the poles of the integrand are at $t=-ks$ for $s=0,1,2,\ldots,\infty$. To evaluate the residue at $-sk$, we put $t=-sk+\theta$, so that the integrand may be reduced to

$$L_1^{ks}\,k^k\frac{1}{\theta^{k-1}}\left(\frac{L_1}{k}\right)^{-\theta}\frac{\Gamma^k(\theta/k+1)}{\Gamma(\theta+1)}\,\frac{\prod\limits_{j=1}^{ks-k+1}(\theta-j)}{\prod\limits_{j=1}^{s}(\theta-jk)^k}. \qquad \ldots\ldots(79)$$

Thus for s greater than or equal to unity, the pole is of order $k-1$ and the residue R_s is given by

$$R_s = (-1)^{k-1}\frac{k^k}{(k-2)!}\frac{(ks-k+1)!}{(s!)^k k^{ks}}L_1^{ks}D(s,k), \qquad \ldots\ldots(80)^*$$

where

$$D(s,k)=\begin{vmatrix} A_1(s) & -1 & 0 & 0 & 0 & \ldots & 0\\ A_2(s) & A_1(s) & -1 & 0 & 0 & \ldots & 0\\ A_3(s) & 2A_2(s) & A_1(s) & -1 & 0 & \ldots & 0\\ A_4(s) & 3A_3(s) & 3A_2(s) & A_1(s) & -1 & \ldots & 0\\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots\\ A_{k-2}(s) & \binom{k-3}{1}A_{k-3}(s) & \binom{k-3}{2}A_{k-4}(s) & \ldots & \ldots & \ldots & A_1(s)\end{vmatrix}, \qquad \ldots\ldots(81)$$

the elements of the determinant being given by the following equations:

$$\left.\begin{aligned} A_1(s) &= \log k-\log L_1-\sum_{j=s+1}^{ks-k+1}\frac{1}{j}\\ A_i(s) &= \left(\frac{1}{k^{i-1}}-1\right)\psi_{i-1}(1)+(-1)^i(i-1)!\left[\sum_{j=1}^{ks-k+1}\frac{1}{j^i}-\frac{1}{k^{i-1}}\sum_{j=1}^{s}\frac{1}{j^i}\right]\\ &\text{for } i=2,3,\ldots, \text{ and } s=1,2,3,\ldots,\infty. \end{aligned}\right\} \qquad \ldots\ldots(82)$$

At $s=0$, the pole is of order k and the corresponding residue is

$$R_0 = \frac{k^k}{(k-2)!\,(k-1)!}D(0,k), \qquad \ldots\ldots(83)$$

where $D(0,k)$ is a determinant similar to the one in equation (81) having $(k-1)$ rows, with $A_i(0)$ for $A_i(s)$, where

$$\left.\begin{aligned} A_1(0) &= \log k-\log L_1-\sum_{j=1}^{k-2}\frac{1}{j}\\ A_i(0) &= \left(\frac{1}{k^{i-1}}-1\right)\psi_{i-1}(1)+(-1)^i(i-1)!\sum_{j=1}^{k-2}\frac{1}{j^i}\quad \text{for } i=2,3,4,\ldots. \end{aligned}\right\} \qquad \ldots\ldots(84)$$

* This is obtained by a method similar to that for (54).

Hence $$F_2(L_1) = \frac{k^k}{(k-2)!}\left[\frac{D(0,k)}{(k-1)!} + (-1)^{k-1}\sum_{s=1}^{\infty}\frac{(ks-k+1)!}{k^{ks}(s!)^k}D(s,k)\,L_1^{ks}\right]. \qquad \ldots\ldots(85)$$

Finally, using equations (76), (77) and (85) we get

$$p(L_1) = \frac{(kn-k-2)!}{\{(n-3)!\}^k\,(k-2)!}\,\frac{L_1^{k(n-2)-1}}{k^{k(n-3)}}$$
$$\times\left[\frac{D(0,k)}{(k-1)!} + (-1)^{k-1}\sum_{s=1}^{\infty}\frac{(ks-k+1)!}{k^{ks}(s!)^k}D(s,k)\,L_1^{ks}\right]. \qquad \ldots\ldots(86)$$

When $k = 2$, the equation (86) takes a simple form:

$$p(L_1) = \frac{(2n-4)!}{(n-3!)^2\,2^{2n-6}}\,L_1^{2n-5}\left[\log\frac{2}{L_1} - \sum_{s=1}^{\infty}\frac{(2s-1)!}{(s!)^2\,2^{2s}}\,L_1^{2s}\right]. \qquad \ldots\ldots(87)$$

The above series may be simplified by integrating both sides of the identity

$$-\frac{1-(1-x^2)^{-\frac{1}{2}}}{x} = \sum_{s=1}^{\infty}\frac{1.3.5\ldots(2s-1)}{s!\,2^s}\,x^{2s-1}.$$

Thus $$-\sum_{s=1}^{\infty}\frac{(2s-1)!}{(s!)^2\,2^{2s}}\,x^{2s} = \int_0^x \frac{1-(1-x^2)^{-\frac{1}{2}}}{x}\,dx$$
$$= -\int_{\sqrt{(1-x^2)}}^{1}\frac{dy}{1+y}, \quad \text{where } 1-x^2 = y^2$$
$$= \log\{1+\sqrt{(1-x^2)}\} - \log 2.$$

It follows therefore that for $k = 2$

$$p(L_1) = \frac{(2n-4)!}{\{(n-3)!\}^2\,2^{2n-6}}\,L_1^{2n-5}\log\frac{1+\sqrt{(1-L_1^2)}}{L_1}, \qquad \ldots\ldots(88)$$

a result given by Pearson & Wilks (1933, p. 367).

5. The independence of the arithmetic mean and the ratio of the geometric to the arithmetic mean of samples drawn from a Pearson type III population

Let $x_1, x_2, \ldots, x_n$ be a sample from the population defined by

$$p(x) = \frac{1}{\Gamma(q)}\,x^{q-1}\,e^{-x}, \quad (0 \leqslant x). \qquad \ldots\ldots(89)$$

Let
$$\bar{x} = (x_1+x_2+\ldots+x_n)/n,$$
$$g = (x_1 x_2 \ldots x_n)^{1/n},$$
$$L = g/\bar{x}. \qquad \ldots\ldots(90)$$

We shall prove that $$p(\bar{x}, L) = p(\bar{x})\,p(L). \qquad \ldots\ldots(91)$$

We start with the simultaneous probability law $p(\bar{x}, g)$ and by transforming the variables to $\bar{x}$ and L, we get

$$p(\bar{x}, L) = p(\bar{x}, g)\,\bar{x}, \qquad \ldots\ldots(92)$$

To find $p(\bar{x}, g)$ consider the function

$$\begin{aligned}
\phi(t, T) &= \int_0^\infty \int_0^\infty e^{-it\bar{x}} g^T p(\bar{x}, g)\, d\bar{x}\, dg \\
&= \frac{1}{[\Gamma(q)]^n} \int_0^\infty \cdots \int_0^\infty e^{-i(x_1+x_2+\ldots+x_n)t/n} (x_1 \ldots x_n)^{T/n} (x_1 x_2 \ldots x_n)^{q-1} \\
&\qquad \times e^{-(x_1+\ldots+x_n)} dx_1 dx_2 \ldots dx_n \\
&= \left[\frac{1}{\Gamma(q)} \int_0^\infty x^{q+T/n-1} e^{-x(1+it/n)} dx\right]^n \\
&= \frac{\Gamma^n(q+T/n)}{\Gamma^n(q)} \left(\frac{n}{n+it}\right)^{nq+T} \qquad \ldots\ldots(93)
\end{aligned}$$

Now we apply theorem 3 and we get

$$\begin{aligned}
p(\bar{x}, g) &= \frac{1}{4\pi^2 i} \int_{-i\infty}^{i\infty} g^{-T-1} dT \int_{-\infty}^{+\infty} e^{it\bar{x}} \phi(t, T)\, dt \\
&= \frac{n^{nq}}{\Gamma^n(q)} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} g^{-T-1} n^T \Gamma^n\left(q + \frac{T}{n}\right) dT \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{it\bar{x}}}{(n+it)^{nq+T}} dt.
\end{aligned}$$

$$\ldots\ldots(94)$$

Since $\displaystyle\int_{-\infty}^{\infty} \frac{e^{ibt}}{(a+it)^x} dt = \frac{e^{-ab} b^{x-1}}{\Gamma(x)}$ for b positive and $R(x) > 0$*

$$p(\bar{x}, g) = \frac{n^{nq}}{\Gamma^n(q)} e^{-n\bar{x}} \bar{x}^{nq-2} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} L^{-T-1} n^T \frac{\Gamma^n\left(q + \frac{T}{n}\right)}{\Gamma(nq+T)} dT.$$

$$\ldots\ldots(95)$$

From equations (92) and (95) we get

$$p(\bar{x}, L) = p(\bar{x}) p(L).$$

6. S. S. Wilks' type B integral equations and the sampling distribution of certain criteria discussed by him

(1) The type B integral equation as defined by Wilks is given by

$$\int_0^\infty z^t f(z)\, dx = \phi(t) = B^t \prod_{j=1}^n \frac{\Gamma(a_j)\Gamma(t+b_j)}{\Gamma(b_j)\Gamma(t+a_j)}, \qquad \ldots\ldots(96)$$

where $a_i \geqslant b_i \geqslant 0$. Let us assume that $\Sigma(a_i - b_i) \geqslant 1$. With this condition it will be seen that $\phi(t)$ belongs to L_2. Hence we may apply theorem 1 and write

$$f(z) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} z^{-t-1} \phi(t)\, dt. \qquad \ldots\ldots(97)$$

Also

$$\phi(t+1) = B \prod_{j=1}^n \frac{t+b_j}{t+a_j} \phi(t). \qquad \ldots\ldots(98)$$

* N. Nielsen (1906), p. 155.

Hence all the conditions of theorem 4 are satisfied and $f(z)$ satisfies the differential equation

$$z \prod_{j=1}^{n} \left(z\frac{d}{dz} - a_j + 2\right) y = B \prod_{j=1}^{n} \left(z\frac{d}{dz} - b_j + 1\right) y. \qquad \ldots\ldots(99)$$

As applications of this principle of solving integral equations, we give below the distribution of the several criteria discussed by Wilks. The notation is the same as that of Wilks and no attempt at defining the criteria is made.

(2) *The generalized correlation ratio, U.*

The tth moment of U is given by Wilks (1932, p. 484) as

$$\phi(t) = \prod_{j=1}^{n} \frac{\Gamma\left(\frac{N-j}{2}\right)}{\Gamma\left(\frac{p-j}{2}\right)} \prod_{j=1}^{n} \frac{\Gamma\left(t+\frac{p-j}{2}\right)}{\Gamma\left(t+\frac{N-j}{2}\right)}. \qquad \ldots\ldots(100)$$

Hence the distribution of U is given by

$$p(U) = \prod_{j=1}^{n} \frac{\Gamma\left(\frac{N-j}{2}\right)}{\Gamma\left(\frac{p-j}{2}\right)} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} U^{-t-1} \prod_{j=1}^{n} \frac{\Gamma\left(t+\frac{p-j}{2}\right)}{\Gamma\left(t+\frac{N-j}{2}\right)} dt. \qquad \ldots\ldots(101)$$

Replacing $t + \frac{1}{2}(p-n)$ by t and putting $N-p = m$,

$$p(U) = \prod_{j=1}^{n} \frac{\Gamma\left(\frac{N-j}{2}\right)}{\Gamma\left(\frac{p-j}{2}\right)} U^{\frac{1}{2}(p-n)-1}$$

$$\times \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} U^{-t} \frac{\Gamma(t)\,\Gamma(t+\frac{1}{2}) \ldots \Gamma\left(t+\frac{n-1}{2}\right)}{\Gamma\left(t+\frac{m}{2}\right)\Gamma\left(t+\frac{m+1}{2}\right) \ldots \Gamma\left(t+\frac{m+n-1}{2}\right)} dt. \qquad \ldots\ldots(102)$$

Denoting the integral on the right side by $F(U)$, $F(U)$ satisfies the differential equation

$$\left(U\frac{d}{dU} - \frac{m-2}{2}\right)\left(U\frac{d}{dU} - \frac{m-1}{2}\right) \ldots \left(U\frac{d}{dU} - \frac{m+n-3}{2}\right) y$$

$$= \frac{d}{dU}\left(U\frac{d}{dU} - \frac{1}{2}\right)\left(U\frac{d}{dU} - 1\right) \ldots \left(U\frac{d}{dU} - \frac{n-1}{2}\right) y. \qquad \ldots\ldots(103)$$

When $n = 1$, (103) reduces to

$$\frac{dy}{dU}(1-U) + \frac{m-2}{2} y = 0, \qquad \ldots\ldots(104)$$

which gives

$$y = C(1-U)^{\frac{1}{2}(m)-1}. \qquad \ldots\ldots(105)$$

* See footnote to p. 284 above regarding use of the letter p.

To find C we compare the residue of $U^{-t}\dfrac{\Gamma(t)}{\Gamma\left(t+\dfrac{m}{2}\right)}$ at $t=0$ with the corresponding term in (105). Thus $C=\dfrac{1}{\Gamma(m/2)}$. Hence

$$p(U)=\frac{\Gamma\left(\dfrac{N-1}{2}\right)U^{\frac{1}{2}(p-3)}}{\Gamma\left(\dfrac{p-1}{2}\right)\Gamma\left(\dfrac{N-p}{2}\right)}(1-U)^{\frac{1}{2}(n-p)-1}. \qquad \text{......(106)}$$

This result has been given by Wilks.

When $n=2$, (103) reduces to

$$U(1-U)\frac{d^2y}{dU^2}+\tfrac{1}{2}\{(2m-5)\,U+1\}\frac{dy}{dU}-\frac{(m-1)\,(m-2)}{4}y=0, \qquad \text{......(107)}$$

from which as in the previous case we get

$$p(U)=\frac{\Gamma(N-2)}{2\Gamma(N-p)\,\Gamma(p-2)}U^{\frac{1}{2}(p-4)}(1-\sqrt{U})^{N-p-1}. \qquad \text{......(108)}$$

Wilks gives this in the form of a hypergeometric series. For $n>2$, $F_2(U)$ cannot be expressed in any simple form, but its value may be obtained as a series.

(3) *Generalization of* $1-\eta^2$, W.

The tth moment of W is given by Wilks (1932, p. 486) as

$$\phi(t)=\prod_{j=1}^{n}\frac{\Gamma\left(\dfrac{N-j}{2}\right)}{\Gamma\left(\dfrac{N-p-j+1}{2}\right)}\prod_{j=1}^{n}\frac{\Gamma\left(\dfrac{N-p-j+1}{2}+t\right)}{\Gamma\left(\dfrac{N-j}{2}+t\right)}. \qquad \text{......(109)}$$

Since (109) may be obtained from (100) by replacing p by $N-p+1$ in the latter, the distribution of W may be readily inferred from that of U.

(4) *Generalization of "Student's" ratio, Y.*

The tth moment of Y (Wilks, 1932, p. 488) is

$$\phi(t)=\frac{\Gamma\left(\dfrac{N}{2}\right)}{\Gamma\left(\dfrac{N-n}{2}\right)}\frac{\Gamma\left(t+\dfrac{N-n}{2}\right)}{\Gamma\left(t+\dfrac{N}{2}\right)}. \qquad \text{......(110)}$$

Hence

$$p(Y)=\frac{\Gamma\left(\dfrac{N}{2}\right)}{\Gamma\left(\dfrac{N-n}{2}\right)}\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}Y^{-t-1}\frac{\Gamma\left(t+\dfrac{N-n}{2}\right)}{\Gamma\left(t+\dfrac{N}{2}\right)}dt. \qquad \text{......(111)}$$

Replacing $t+\frac{1}{2}(N-n)$ by t,

$$p(Y) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)} Y^{\frac{1}{2}(N-n)-1} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} Y^{-t} \frac{\Gamma(t)}{\Gamma\left(t+\frac{n}{2}\right)} dt. \qquad \text{......(112)}$$

It may be proved that the integral in (112) satisfies the differential equation

$$\left(Y\frac{d}{dY}-\frac{n}{2}\right) Yz = Y\frac{dz}{dY}. \qquad \text{......(113)}$$

Solving this and finding the constants so as to make the solution identical to $p(Y)$, we get

$$p(Y) = \frac{\Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{n}{2}\right)} Y^{\frac{1}{2}(N-n)-1}(1-Y)^{\frac{1}{2}(n)-1}. \qquad \text{......(114)}$$

This result is given by Wilks.

(5) *Ratios of determinants of correlation coefficients, w.*

The tth moment of w is given by Wilks (1932, p. 491) as

$$\phi(t) = \frac{\Gamma^{n-1}\left(\frac{N-1}{2}\right) \prod_{j=2}^{n} \Gamma\left(\frac{N-j}{2}+t\right)}{\prod_{j=2}^{n} \Gamma\left(\frac{N-j}{2}\right) \Gamma^{n-1}\left(\frac{N-1}{2}+t\right)}. \qquad \text{......(115)}$$

Hence
$$p(w) = \frac{\Gamma^{n-1}\left(\frac{N-1}{2}\right)}{\prod_{j=2}^{n} \Gamma\left(\frac{N-j}{2}\right)} \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} w^{-t-1} \frac{\prod_{j=2}^{n} \Gamma\left(\frac{N-j}{2}+t\right)}{\Gamma^{n-1}\left(\frac{N-1}{2}+t\right)} dt. \qquad \text{......(116)}$$

Putting t instead of $t+\frac{1}{2}(N-n)$ we get

$$p(w) = \frac{\Gamma^{n-1}\left(\frac{N-1}{2}\right)}{\prod_{j=2}^{n} \Gamma\left(\frac{N-j}{2}\right)} w^{\frac{1}{2}(N-n)-1} \frac{1}{2\pi i} \int_{-i\infty+\frac{1}{2}(N-n)}^{i\infty+\frac{1}{2}(N-n)} w^{-t} \frac{\prod_{j=2}^{n} \Gamma\left(\frac{n-j}{2}+t\right)}{\Gamma^{n-1}\left(\frac{n-1}{2}+t\right)} dt. \qquad \text{......(117)}$$

The integral in (117) satisfies the differential equation

$$\frac{d}{dw}\left(w\frac{d}{dw}-\frac{1}{2}\right)\left(w\frac{d}{dw}-1\right)\ldots\left(w\frac{d}{dw}-\frac{n-2}{2}\right) y = \left(w\frac{d}{dw}-\frac{n-3}{2}\right)^{n-1} y. \qquad \text{......(118)}$$

For $n = 2$, 3 this equation is readily solved, giving respectively

$$p(w) = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\sqrt{\pi}} w^{\frac{1}{2}(N-2)-1}(1-w)^{-\frac{1}{2}} \quad (n = 2), \qquad \ldots\ldots(119)$$

and $$p(w) = \sqrt{\pi}\frac{\Gamma^2\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-2}{2}\right)\Gamma\left(\frac{N-3}{2}\right)} w^{\frac{1}{2}(N-3)-1}\left[1-\frac{2}{\pi}\sin^{-1}\sqrt{w}\right] \quad (n = 3). \qquad \ldots\ldots(120)$$

Equation (120) may be compared with the one given by Wilks in the paper cited (1932, p. 492, 49*b*).

7. Conclusion

The main object of the paper has been to develop the study of distribution laws of statistics whose moment function can be evaluated. The method gives an elegant mathematical solution to Wilks' type B integral equation. A detailed study of these functions from the mathematical point of view is made possible since the differential equation that they satisfy can readily be written down.

In conclusion, the author wishes to thank Professor E. S. Pearson under whose suggestion and guidance the present paper was written. The author is also indebted to Dr G. Rasch for considerable help he has had in developing the ideas of § 2, and to Dr R. C. Geary for a number of suggestions which have improved the final form of the paper.

8. Appendix. Determinants arising out of the successive derivatives of exponential functions

(1) Let (a_i) be a sequence of numbers, finite or infinite, and let

$$F(x;\, t;\, a_2, a_3, \ldots) \equiv e^{xt+a_2\frac{t^2}{2!}+a_3\frac{t^3}{3!}+\ldots\infty}. \qquad \ldots\ldots(1)$$

Then the nth derivative of $F(x;\, t;\, a_2, a_3, \ldots)$ at $t = 0$ is

$$D_n(x,a) = \begin{vmatrix} x & -1 & 0 & 0 & 0 & \ldots & 0 \\ a_2 & x & -1 & 0 & \ldots & \ldots & 0 \\ a_3 & 2a_2 & x & -1 & \ldots & \ldots & 0 \\ a_4 & 3a_3 & 3a_2 & x & -1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\ a_n & \binom{n-1}{1}a_{n-1} & \binom{n-1}{2}a_{n-2} & \ldots & \ldots & \ldots & x \end{vmatrix} \qquad \ldots\ldots(2)$$

To prove this, we differentiate both sides of (1), so that

$$F_1(x;\, t;\, a) = F(x;\, t;\, a)\left[x + a_2 t + a_3 \frac{t^2}{2!} + \ldots\right]. \qquad \ldots\ldots(3)$$

Using Leibnitz' rule on successive derivatives to (3)

$$F_n(x;\, t;\, a) = \sum_{i=0}^{n-1} \binom{n-1}{i} F_{n-i-1}(x;\, t;\, a)\,[a_{i+1} + a_{i+2} t + \ldots]. \qquad \ldots\ldots(4)$$

Putting $t = 0$ in (4) and giving n the values $1, 2, 3, \ldots, n$ and eliminating $F_1, F_2, \ldots, F_{n-1}$, we get (2).

(2) If $c_0, c_1, c_2, \ldots, c_n$ are constants such that

$$c_0 + c_1 x + c_2 x^2 + \ldots + c_n x^n = D_n(x, b), \qquad \ldots\ldots(5)$$

then

(1) $c_0 + c_1 D_1(x, a) + c_2 D_2(x, a) + \ldots + c_n D_n(x, a) = D_n(x, a+b)$. $\qquad \ldots\ldots(6)$

(2) $c_0 - c_1 D_1(x, a) + c_2 D_2(x, a) + \ldots + (-1)^n c_n D_n(x, a) = (-1)^n D_n(x, a_i + (-)^i b_i)$. $\qquad \ldots\ldots(7)$

To prove (6), we note that

$$D_n(x, b) = F(x;\, 0;\, b) = \left(\frac{d}{dt}\right)^n_{t=0} F(x;\, t;\, b).$$

Hence from (1),

$$\sum_{i=0}^{n} c_i x^i = \left(\frac{d}{dt}\right)^n_{t=0} F(x;\, t;\, b). \qquad \ldots\ldots(8)$$

Using (8), we write (6) as

$$\begin{aligned}
&\sum_{i=0}^{n} c_i \left(\frac{d}{dt}\right)^i F(x;\, t;\, a)\bigg|_{t=0} \\
&= \left[\sum_{i=0}^{n} c_i \left(\frac{d}{dt}\right)^i\right] F(x;\, t;\, a)\bigg|_{t=0} \\
&= \left[\left(\frac{d}{d\theta}\right)^n F\left(\frac{d}{dt};\, \theta;\, b\right)\right]_{\theta=0} F(x;\, t;\, a)\bigg|_{t=0} \\
&= \left(\frac{d}{d\theta}\right)^n e^{\theta \frac{d}{dt}} F(0;\, \theta;\, b)\, F(x;\, t;\, a)\bigg|_{t=\theta=0} \\
&= e^{\theta \frac{d}{dt}} \left(\frac{d}{d\theta} + \frac{d}{dt}\right)^n F(0;\, \theta;\, b)\, F(x;\, t;\, a)\bigg|_{t=\theta=0} \\
&= \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{d\theta}\right)^i F(0;\, \theta;\, b) \left(\frac{d}{dt}\right)^{n-i} F(x;\, t;\, a)\bigg|_{t=\theta=0} \\
&= \sum_{i=0}^{n} \binom{n}{i} \left(\frac{d}{dt}\right)^i F(0;\, t;\, b) \left(\frac{d}{dt}\right)^{n-i} F(x;\, t;\, a)\bigg|_{t=0} \\
&= \left(\frac{d}{dt}\right)^n F(0;\, t;\, b)\, F(x;\, t;\, a)\bigg|_{t=0} \\
&= \left(\frac{d}{dt}\right)^n F(x;\, t;\, a+b)\bigg|_{t=0} \\
&= D_n(x;\, a+b).
\end{aligned}$$

To prove (7), we have

$$\sum_{i=0}^{n}(-1)^i c_i D_i(x,a) = \sum_{i=0}^{n}(-1)^i c_i\left(\frac{d}{dt}\right)^i F(x;t;a)\bigg|_{t=0}$$

$$= \left[\sum_{i=0}^{n} c_i\left(-\frac{d}{dt}\right)^i\right] F(x;t;a)\bigg|_{t=0}$$

$$= \left[\left(\frac{d}{d\theta}\right)^n F\left(-\frac{d}{dt};\theta;b\right)\right]_{\theta=0} F(x;t;a)\bigg|_{t=0}$$

$$= \left(\frac{d}{d\theta}\right)^n e^{-\theta\frac{d}{dt}} F(0;\theta;b)\, F(x;t;a)\bigg|_{t=\theta=0}$$

$$= (-1)^n\left(\frac{d}{d\theta}\right)^n e^{\theta\frac{d}{dt}} F(0;-\theta;b)\, F(x;t;a)\bigg|_{t=\theta=0}$$

$$= (-1)^n\left(\frac{d}{d\theta}+\frac{d}{dt}\right)^n F(0;-\theta;b)\, F(x;t;a)\bigg|_{t=\theta=0}$$

$$= (-1)^n\sum_{i=0}^{n}\binom{n}{i}\left(\frac{d}{d\theta}\right)^i F(0;-\theta;b)\left(\frac{d}{dt}\right)^{n-i} F(x;t;a)\bigg|_{t=\theta=0}$$

$$= (-1)^n\sum_{i=0}^{n}\binom{n}{i}\left(\frac{d}{dt}\right)^i F(0;-t;b)\left(\frac{d}{dt}\right)^{n-i} F(x;t;a)\bigg|_{t=0}$$

$$= (-1)^n\left(\frac{d}{dt}\right)^n F(0;-t;b)\, F(x;t;a)\bigg|_{t=0}$$

$$= (-1)^n\left(\frac{d}{dt}\right)^n F(x;t,a_i+(-1)^i b_i)$$

$$= (-1)^n D_n(x;a_i+(-1)^i b_i).$$

REFERENCES

Bartlett, M. S. (1937*a*). *Proc. Roy. Soc.* A, **160**, 268–82.

—— (1937*b*). *J.R. Statist. Soc. Suppl.* **4**, 137–70 (158).

Forsyth, A. R. (1921). *A Treatise on Differential Equations*.

Nayer, P. P. N. (1936). *Statist. Res. Mem.* **1**, 38–51.

Neyman, J. & Pearson, E. S. (1931). *Bull. int. Acad. Cracovie*, A, pp. 460–8.

Nielsen, N. (1906). *Handbuch der Gamma Funktion*.

Pearson, E. S. & Wilks, S. S. (1933). *Biometrika*, **25**, 353–78.

Wilks, S. S. (1932). *Biometrika*, **24**, 471–94.

SAMPLING DISTRIBUTION AND SELECTION IN A NORMAL POPULATION

By WALTER LEDERMANN
Moray House, University of Edinburgh

CONTENTS

1. Introduction

We consider a normal population which is specified by two sets of p and q variates respectively, whose variances and covariances are arranged in a matrix

$$R = \begin{bmatrix} R_{pp} & R_{pq} \\ R_{qp} & R_{qq} \end{bmatrix}, \qquad \ldots\ldots(1)$$

which hereafter will be called the *variance matrix* of the $p+q$ variates. To distinguish the two sets of variates we have written the matrix R in a partitioned form; thus R_{pp} is the variance matrix of the first set and R_{qq} that of the second set, while R_{pq} contains as elements the pq covariances between any variate of the first set and any variate of the second set. We have also put

$$(R_{pq})' = R_{qp}, \qquad \ldots\ldots(2)$$

a convention to which we shall adhere in the case of all rectangular matrices whose orders are indicated by the suffixes p and q.

Suppose now that a selection is carried out in the population in such a way that (i) all variates remain normally distributed, and (ii) that the variance matrix of the first set is changed from R_{pp} to V_{pp} which may be any preassigned matrix, provided it is symmetrical and positive definite. Owing to the statistical dependence between the $p+q$ characters, the other variances and covariances will also be modified, and it is known that the variance matrix after selection is given by

$$V = \begin{bmatrix} V_{pp} & V_{pq} \\ V_{qp} & V_{qq} \end{bmatrix} = \left[\begin{array}{c:c} V_{pp} & V_{pp} R_{pp}^{-1} R_{pq} \\ R_{qp} R_{pp}^{-1} V_{pp} & R_{qq} - R_{qp}(R_{pp}^{-1} - R_{pp}^{-1} V_{pp} R_{pp}^{-1}) R_{pq} \end{array}\right]. \qquad \ldots\ldots(3)$$

This problem was first solved by K. Pearson (1903). The matrix form in which we have quoted his result is due to Aitken (1935, 1936). The above formula can be

obtained without any reference to the statistical method by which the change of the variances and covariances in the first set is effected.

It is the object of this paper to show that selection can be regarded as the limiting case of a certain regression problem with respect to the population of variance matrices computed for all possible samples of n individuals: suppose that the variance matrix for an arbitrary sample of n persons is

$$Z = \begin{bmatrix} Z_{pp} & Z_{pq} \\ Z_{qp} & Z_{qq} \end{bmatrix}.$$

The matrix Z will, of course, vary from one sample to another and will also depend on n, the number of persons in the sample; in other words we shall obtain a population of matrices Z which will possess a certain distribution law (see §5 below). Consider now the subpopulation or "array" of those matrices Z in which the first submatrix Z_{pp} is equal to a given matrix V_{pp}. Our task will then be to find the mean value or "expected" value V^* of this array. Evidently V^* will be a function of V_{pp}. The chief result is that the mean V^* of this subpopulation of Z-matrices tends to the matrix V (equation (3)) as n tends to infinity. Thus *selection in Pearson's sense means finding the average value of the variance matrix with respect to the population of all possible* infinite *normal samples which are subject to the condition that the variance matrix of the first set of variates is equal to the preassigned matrix* V_{pp}.

This idea was communicated to the present writer by Prof. Godfrey H. Thomson, who has discussed some of the consequences elsewhere (1939). In this paper† we propose to give an analytical proof of Prof. Thomson's statement by deriving an explicit formula for the average of the variance matrix under the conditions referred to.

2. Moment generating function for an array

Consider two sets of variates

$$x = \{x_1, x_2, \ldots, x_p\}$$

and

$$y = \{y_1, y_2, \ldots, y_q\}.$$

which are envisaged as column vectors of orders p and q respectively, and suppose that their frequency differential is given by

$$\phi(x, y)\, dx dy,$$

where dx stands for $dx_1, dx_2, \ldots, dx_p$, and dy for $dy_1, dy_2, \ldots, dy_q$. The moment generating function of x and y is then

$$g(t, s) = \iint_{-\infty}^{\infty} \phi(x, y)\, e^{i(t'x+s'y)}\, dx dy, \qquad \ldots\ldots(\mathrm{G})$$

† The author wishes to express his thanks to Prof. Godfrey H. Thomson for suggesting this problem to him. He is also indebted to a referee for making some valuable criticisms, especially in connexion with the subject of § 4 below.

where the vectors t and s are the moment carrying variables representing x and y respectively; the accent, as usual, denotes the transposition of a matrix, so that t' is a row vector and

$$t'x = \sum_{i=1}^{p} t_i x_i,$$

and similarly

$$s'y = \sum_{i=1}^{q} s_i y_i.$$

We now consider an x-array of the variates, i.e. we assign some constant values η to the variates y. The distribution function of the remaining variates x is then evidently given by

$$\phi^*(x) = \frac{\phi(x, \eta)}{\int_{-\infty}^{\infty} \phi(x, \eta)\, dx} = c\phi(x, \eta),$$

and the corresponding moment generating function becomes

$$g^*(t) = c\int_{-\infty}^{\infty} \phi(x, \eta)\, e^{it'x}\, dx. \qquad \text{......(G*)}$$

On the other hand, using the Fourier integral theorem on equation (G) we find

$$\int_{-\infty}^{\infty} g(t, s)\, e^{-is'\eta}\, ds = (2\pi)^q \int_{-\infty}^{\infty} \phi(x, \eta)\, e^{it'x}\, dx,$$

whence, comparing the last equation with (G*), we obtain the result

$$g^*(t) = \text{const.} \int_{-\infty}^{\infty} g(t, s)\, e^{-is'\eta}\, ds. \qquad \text{......(4)}$$

When working with moment generating functions it should be borne in mind that the constant term, i.e. the term independent of the moment carrying symbols, is always equal to unity. Hence throughout the analysis we can neglect any non-zero multiplicative constant; and in the final result we can restore the correct constant by making the first term in the expansion of the moment generating function equal to unity.

3. SOME LEMMAS ON MATRICES AND DETERMINANTS

(i) Let

$$S = \begin{bmatrix} S_{pp} & S_{pq} \\ S_{qp} & S_{qq} \end{bmatrix}$$

be any square matrix which is partitioned as shown, the suffixes indicating the number of rows and columns for each of the four submatrices, and suppose that $|S_{qq}| \neq 0$. It is then easy to verify the matrix identity

$$\begin{bmatrix} S_{pp} & S_{pq} \\ S_{qp} & S_{qq} \end{bmatrix} \begin{bmatrix} I & O \\ -S_{qq}^{-1} S_{qp} & I \end{bmatrix} = \begin{bmatrix} S_{pp} - S_{pq} S_{qq}^{-1} S_{qp} & S_{pq} \\ O & S_{qq} \end{bmatrix},$$

whence on taking determinants

$$|S| = |S_{qq}| \times |S_{pp} - S_{pq} S_{qq}^{-1} S_{qp}|. \qquad \text{......(5)}$$

(ii) If $X=[x_{ik}]$ be any matrix, we shall use the symbol $O(x^2)$ to denote any function (scalar or matrix function) of the x_{ik} which, when expanded as a power series, involves only terms which are at least of the second order in the x_{ik}. E.g. for an arbitrary square matrix X we have

$$|I+X| = 1+\operatorname{tr} X + O(x^2), \qquad \text{......(6)}$$

where

$$\operatorname{tr} X = \sum_i x_{ii}$$

denotes the "trace" of X.

(iii) We shall frequently use the relations

$$\left.\begin{aligned}\operatorname{tr}\{AB\} &= \operatorname{tr}\{BA\},\\ \operatorname{tr}\{ABC\} &= \operatorname{tr}\{BCA\} = \operatorname{tr}\{CAB\}\end{aligned}\right\}, \qquad \text{......(7)}$$

the general rule being that the trace of a product of matrices is unaltered when the factors are permuted in cyclical order.

4. Ingham's integral

Let $U=[u_{\alpha\beta}]$ and $V=[v_{\alpha\beta}]$ be any given positive definite matrices of order p, and let $T=[t_{\alpha\beta}]$ be a variable symmetrical matrix whose $\frac{1}{2}p(p+1)$ distinct elements are regarded as independent variables. Then A. E. Ingham has proved (1933) that

$$\left(\frac{1}{2\pi}\right)^{\frac{1}{2}p(p+1)}\int_{-\infty}^{\infty}|U-iT|^{-h}e^{-i\operatorname{tr}(TV)}\,dT = e^{-\operatorname{tr}(UV)}\,J(V,h), \qquad \text{......(8)}$$

where

$$J(V,h) = (2\sqrt{\pi})^{-\frac{1}{2}p(p-1)}\,|V|^{h-\frac{1}{2}(p+1)}\prod_{\mu=0}^{p-1}\left\{\Gamma\left(h-\frac{\mu}{2}\right)\right\}^{-1}. \qquad \text{......(9)}$$

The integral (8) is an $\frac{1}{2}p(p+1)$-fold integral to be extended over the $\frac{1}{2}p(p+1)$ distinct elements of the symmetrical matrix T; accordingly we have introduced the abbreviation

$$dT = dt_{11}dt_{12}\ldots dt_{pp} = \prod_{\alpha\leqslant\beta}{}_1^p\, dt_{\alpha\beta}.$$

Ingham has shown that the integral converges absolutely when $h>\frac{1}{2}p(p+1)$. This condition will in general be fulfilled in our problem, because h will be identified with $\frac{1}{2}(n-1)$ where n is the number of persons in a sample, and p will be the number of directly selected tests.

Further, we shall need for our purpose to extend the validity of (8) to the case where the matrix U has complex numbers as its elements, provided that the real parts of the elements form a positive definite matrix. This is easily done. Suppose that

$$U = U_1+iU_2,$$

where U_1 and U_2 are real symmetric matrices and where U_1 is positive definite. We have then for the left-hand side of (8), the integral

$$\left(\frac{1}{2\pi}\right)^{\frac{1}{2}p(p+1)}\int_{-\infty}^{\infty}|U_1-i(T-U_2)|^{-h}e^{-i\operatorname{tr}(TV)}\,dT.$$

A change of origin $$W = T - U_2,$$
where W is the matrix of the new variables $w_{11}, w_{12}, \ldots, w_{pp}$, gives us

$$\left(\frac{1}{2\pi}\right)^{\frac{1}{2}p(p+1)} e^{-i\,\mathrm{tr}(U_2V)} \int_{-\infty}^{\infty} |\, U_1 - iW\,|^{-h} e^{-i\,\mathrm{tr}(WV)}\, dW.$$

In the last integral U_1 is real and positive definite, so according to (8), we obtain the result
$$e^{-i\,\mathrm{tr}(U_2V)}\, e^{-\mathrm{tr}(U_1V)}\, J(V, h),$$
i.e.
$$e^{-\mathrm{tr}(UV)}\, J(V, h),$$
which is exactly the same as the right-hand side of (8). The desired extension is therefore achieved.

For our purpose it is sufficient to know that the expression (9) is independent of U, and we shall write the result in the form

$$\int_{-\infty}^{\infty} |\, U - iT\,|^{-h} e^{-i\,\mathrm{tr}(TV)}\, dT = \text{const.}\, e^{-\mathrm{tr}(UV)}, \qquad \ldots\ldots(10)$$

the elements of the matrix V being treated as constants.

5. SAMPLING DISTRIBUTION OF VARIANCES AND COVARIANCES

We consider all possible samples of n individuals drawn from a $(p+q)$-variate normal population. Each sample will have its own variance matrix

$$Z = \begin{bmatrix} Z_{pp} & Z_{pq} \\ Z_{qp} & Z_{qq} \end{bmatrix}.$$

As we pass from one sample to another the matrices Z will form a population whose distribution function has been worked out (Wishart & Bartlett, 1933). The moment generating function of this distribution can be written in the form (*loc. cit.* p. 269)

$$g(T) = |\, R\,|^{-\frac{1}{2}(n-1)}\, |\, R^{-1} - \frac{2i}{n} T\,|^{-\frac{1}{2}(n-1)}, \qquad \ldots\ldots(11)$$

where R is the variance matrix of the original population. The symbols $it_{\alpha\alpha}$ and $2it_{\alpha\beta}\,(\alpha<\beta)$ are the moment carrying variables for the variances $z_{\alpha\alpha}$ and the covariances $z_{\alpha\beta}\,(\alpha<\beta)$ respectively. Thus, if the expansion of $g(T)$ as far as linear terms in t be
$$g(T) = 1 + i \sum_{\alpha} w_{\alpha\alpha} t_{\alpha\alpha} + 2i \sum_{\alpha<\beta} w_{\alpha\beta} t_{\alpha\beta} + O(t^2),$$
it would follow that the mean value of $z_{\alpha\alpha}$ is $w_{\alpha\alpha}$, and that the mean value of $t_{\alpha\beta}$ is $w_{\alpha\beta}$. The last equation can be more conveniently written in the form

$$g(T) = 1 + i\,\mathrm{tr}\{TW'\} + O(t^2). \qquad \ldots\ldots(12)$$

Incidentally, with this notation it is quite easy to deduce the well-known result

$$W = \frac{n-1}{n} R.$$

For we can rewrite (11) thus

$$g(T) = \left| I - \frac{2i}{n} TR \right|^{-\frac{1}{2}(n-1)} = \left\{ 1 - \frac{2i}{n} \operatorname{tr}(TR) + O(t^2) \right\}^{-\frac{1}{2}(n-1)},$$

by (6), p. 298. Expanding the expression on the right-hand side we obtain

$$g(T) = 1 + \frac{n-1}{n} i \operatorname{tr}(TR) + O(t^2).$$

Therefore, by comparing this with (12)

$$W' = W = \frac{n-1}{n} R.$$

6. Moment generating function for an array $\tilde{Z}_{pp} = V_{pp}$

We now consider the array of the Z-distribution in which the variables Z_{pp} have assigned values† Z_{pp}. According to the results of § 2, the moment generating function $g^*(T)$ for this subpopulation of Z-matrices is obtained from $g(T)$ by applying a Fourier transformation with respect to those variables which are kept constant. Thus

$$g^*(T) \propto \int_{-\infty}^{\infty} \left| R^{-1} - \frac{2i}{n} T \right|^{-\frac{1}{2}(n-1)} e^{-i \operatorname{tr}(T_{pp} V_{pp})} dT_{pp}. \qquad \ldots\ldots(13)$$

The integration refers to the $\frac{1}{2}p(p+1)$ elements of the symmetrical matrix T_{pp}; we have suppressed the normalizing factor of the integral and the constant factor $|R|^{-\frac{1}{2}(n-1)}$ of the function $g(T)$. In order to evaluate the integral we temporarily put

$$R^{-1} = \begin{bmatrix} Q_{pp} & Q_{pq} \\ Q_{qp} & Q_{qq} \end{bmatrix}, \qquad \ldots\ldots(14)$$

and

$$S = R^{-1} - \frac{2i}{n} T = \begin{bmatrix} S_{pp} & S_{pq} \\ S_{qp} & S_{qq} \end{bmatrix} = \begin{bmatrix} Q_{pp} - \frac{2i}{n} T_{pp} & Q_{pq} - \frac{2i}{n} T_{pq} \\ Q_{qp} - \frac{2i}{n} T_{qp} & Q_{qq} - \frac{2i}{n} T_{qq} \end{bmatrix}. \qquad \ldots\ldots(15)$$

Hence by (5), p. 297,

$$\left| R^{-1} - \frac{2i}{n} T \right| = |S| = |S_{qq}| \times |S_{pp} - S_{pq} S_{qq}^{-1} S_{qp}|.$$

Substituting this in (13) and noting that S_{qq} is independent of the variables of integration we find

$$g^*(T) \propto |S_{qq}|^{-\frac{1}{2}(n-1)} \int_{-\infty}^{\infty} |S_{pp} - S_{pq} S_{qq}^{-1} S_{qp}|^{-\frac{1}{2}(n-1)} e^{-i \operatorname{tr}(T_{pp} V_{pp})} dT_{pp}.$$

Now let the matrix U_{pp} be defined such that

$$\frac{2}{n} U_{pp} = Q_{pp} - S_{pq} S_{qq}^{-1} S_{qp}, \qquad \ldots\ldots(16)$$

† The matrix V_{pp} is, of course, symmetrical and positive definite.

which is constant with respect to the integration. Then

$$|S_{pp}-S_{pq}S_{qq}^{-1}S_{qp}| = \left|\frac{2}{n}U_{pp}-\frac{2i}{n}T_{pp}\right| = \left(\frac{2}{n}\right)^p |U_{pp}-iT_{pp}|,$$

and the moment generating function of the array becomes

$$g^*(T) \propto |S_{qq}|^{-\frac{1}{2}(n-1)} \int_{-\infty}^{\infty} |U_{pp}-iT_{pp}|^{-\frac{1}{2}(n-1)} e^{-i\,\mathrm{tr}(T_{pp}V_{pp})}\,dT_{pp},$$

where numerical factors have been ignored. The integral on the right-hand side of the last equation is precisely of the type discussed by Ingham, whence by (10), p. 299, we can write

$$g^*(T) \propto |S_{qq}|^{-\frac{1}{2}(n-1)} e^{-\mathrm{tr}(U_{pp}V_{pp})}. \qquad \text{......(17)}$$

On the other hand, the expansion of $g^*(T)$ must be of the form

$$g^*(T) = 1 + 2i\,\mathrm{tr}(T_{pq}V_{qp}^*) + i\,\mathrm{tr}(T_{qq}V_{qq}^*) + O(t^2), \qquad \text{......(18)}$$

there being no term in T_{pp} since the corresponding variables are now fixed. Our object in the next section will be to expand (17) as far as linear terms in the $t_{\alpha\beta}$; a comparison with (18) will then immediately yield the mean values of the variables Z_{pq} and Z_{qq} in the array $Z_{pp} = V_{pp}$.

In order to justify the application of the extended form of Ingham's integral in our case we still have to show that the real part of the matrix U_{pp} defined in equation (16) is positive definite. Denoting this matrix by $U_{pp}^{(1)}$, it is seen that the elements of $U_{pp}^{(1)}$ are real continuous functions of the elements of the matrices T_{pq} and T_{qq}. At the point $T_{pq} = 0$ and $T_{qq} = 0$, the matrix $U_{pp}^{(1)}$ is reduced to, say, $U_{pp}^{(0)}$, where

$$\frac{2}{n}U_{pp}^{(0)} = Q_{pp} - Q_{pq}Q_{qq}^{-1}Q_{qp}, \qquad \text{......(19)}$$

in accordance with (15) and (16). It is sufficient for our purpose to show that $U_{pp}^{(0)}$ is positive definite. For then, by continuity, $U_{pp}^{(1)}$ will remain positive at least for a certain range of values $T_{pq} \neq 0$ and $T_{qq} \neq 0$, and in the expansion (18) the independent variables may be restricted to as small a range as we please.

In order to show the positive definiteness of $U_{pp}^{(0)}$, we express the right-hand side of (19) in terms of the matrix R as follows: by (14) we have

$$I = RQ,$$

i.e.

$$\begin{bmatrix} I & O \\ O & I \end{bmatrix} = \begin{bmatrix} R_{pp} & R_{pq} \\ R_{qp} & R_{qq} \end{bmatrix} \begin{bmatrix} Q_{pp} & Q_{pq} \\ Q_{qp} & Q_{qq} \end{bmatrix} = \begin{bmatrix} R_{pp}Q_{pp}+R_{pq}Q_{qp} & R_{pp}Q_{pq}+R_{pq}Q_{qq} \\ R_{qp}Q_{pp}+R_{qq}Q_{qp} & R_{qp}Q_{pq}+R_{qq}Q_{qq} \end{bmatrix}$$

Hence $$O = R_{pp}Q_{pq} + R_{pq}Q_{qq},$$

or $$R_{pp}^{-1}R_{pq} = -Q_{pq}Q_{qq}^{-1}, \qquad \text{......(20)}$$

and $$I = R_{pp}Q_{pp} + R_{pq}Q_{qp},$$

$$R_{pp}^{-1} = Q_{pp} + R_{pp}^{-1}R_{pq}Q_{qp},$$

whence by (20) $$R_{pp}^{-1} = Q_{pp} - Q_{pq}Q_{qq}^{-1}Q_{qp}. \qquad \text{......(21)}$$

Similarly, we can deduce the identity

$$Q_{qq}^{-1} = R_{qq} - R_{qp} R_{pp}^{-1} R_{pq}. \qquad \text{......(22)}$$

On substituting (21) in (19) we see that

$$U_{pp}^{(o)} = \tfrac{1}{2} n R_{pp}^{-1}.$$

But since R is a positive definite matrix, so is R_{pp} and R_{pp}^{-1}, and consequently also $U_{pp}^{(o)}$.

7. The linear terms of the moment generating function

In order to find the linear terms of the function $g^*(T)$ (equation (17), p. 301) we write

$$g^*(T) \propto f_1 . f_2,$$

where

$$f_1 = |S_{qq}|^{-\frac{1}{2}(n-1)}$$

and

$$f_2 = e^{-\operatorname{tr}(U_{pp} V_{pp})},$$

and expand each factor separately. First we have, by (15),

$$|S_{qq}|^{-\frac{1}{2}(n-1)} = \left| Q_{qq} - \frac{2i}{n} T_{qq} \right|^{-\frac{1}{2}(n-1)} = |Q_{qq}|^{-\frac{1}{2}(n-1)} \left| I - \frac{2i}{n} Q_{qq}^{-1} T_{qq} \right|^{-\frac{1}{2}(n-1)},$$

$$|S_{qq}|^{-\frac{1}{2}(n-1)} = |Q_{qq}|^{-\frac{1}{2}(n-1)} \left\{ 1 - \frac{2i}{n} \operatorname{tr}(Q_{qq}^{-1} T_{qq}) + O(t^2) \right\}^{-\frac{1}{2}(n-1)},$$

$$|S_{qq}|^{-\frac{1}{2}(n-1)} = |Q_{qq}|^{-\frac{1}{2}(n-1)} \left\{ 1 + \frac{n-1}{n} i \operatorname{tr}(Q_{qq}^{-1} T_{qq}) + O(t^2) \right\},$$

by an argument similar to that used at the end of § 5, thus

$$f_1 \propto 1 + \frac{n-1}{n} i \operatorname{tr}(T_{qq} Q_{qq}^{-1}) + O(t^2). \qquad \text{......(23)}$$

Next, consider

$$f_2 = e^{-\operatorname{tr}(U_{pp} V_{pp})},$$

where, by (15) and (16)

$$\frac{2}{n} U_{pp} = Q_{pp} - \left(Q_{pq} - \frac{2i}{n} T_{pq} \right) \left(Q_{qq} - \frac{2i}{n} T_{qq} \right)^{-1} \left(Q_{qp} - \frac{2i}{n} T_{qp} \right).$$

But

$$\left(Q_{qq} - \frac{2i}{n} T_{qq} \right)^{-1} = \left[Q_{qq} \left(I - \frac{2i}{n} Q_{qq}^{-1} T_{qq} \right) \right]^{-1} = \left(I - \frac{2i}{n} Q_{qq}^{-1} T_{qq} \right)^{-1} Q_{qq}^{-1},$$

$$\left(Q_{qq} - \frac{2i}{n} T_{qq} \right)^{-1} = \left(I + \frac{2i}{n} Q_{qq}^{-1} T_{qq} + O(t^2) \right) Q_{qq}^{-1},$$

since for matrices with sufficiently small elements

$$(I - X)^{-1} = I + X + O(x^2).$$

Hence

$$\left(Q_{qq} - \frac{2i}{n} T_{qq} \right)^{-1} = Q_{qq}^{-1} + \frac{2i}{n} Q_{qq}^{-1} T_{qq} Q_{qq}^{-1} + O(t^2),$$

and the expression for U_{pp} becomes

$$\frac{2}{n} U_{pp} = Q_{pp} - \left(Q_{pq} - \frac{2i}{n} T_{pq} \right) \left(Q_{qq}^{-1} + \frac{2i}{n} Q_{qq}^{-1} T_{qq} Q_{qq}^{-1} \right) \left(Q_{qp} - \frac{2i}{n} T_{qp} \right) + O(t^2).$$

After expanding and collecting terms which are linear in the $t_{\alpha\beta}$ we obtain

$$\frac{2}{n} U_{pp} V_{pp} = C + \frac{2i}{n} \{T_{pq} Q_{qq}^{-1} Q_{qp} - Q_{pq} Q_{qq}^{-1} T_{qq} Q_{qq}^{-1} Q_{qp} + Q_{pq} Q_{qq}^{-1} T_{qp}\} V_{pp} + O(t^2),$$

where C is a certain constant matrix whose trace we shall denote by c. When taking the trace of each side in the last equation we shall rearrange the factors in every term in such a way that T_{pq} or T_{qq} occupies the first place and T_{qp} occupies the last place. This can always be done by a cyclical permutation of the factors. Thus

$$\frac{2}{n} \operatorname{tr}(U_{pp} V_{pp}) = c + \frac{2i}{n} [\operatorname{tr}(T_{pq} Q_{qq}^{-1} Q_{qp} V_{pp}) - \operatorname{tr}(T_{qq} Q_{qq}^{-1} Q_{qp} V_{pp} Q_{pq} Q_{qq}^{-1}) + \operatorname{tr}(V_{pp} Q_{pq} Q_{qq}^{-1} T_{qp})] + O(t^2).$$

Now the third term in the square bracket is equal to the first term, since the trace of a matrix is equal to that of its transpose. Hence

$$\operatorname{tr}(U_{pp} V_{pp}) = \tfrac{1}{2} nc + 2i \operatorname{tr}(T_{pq} Q_{qq}^{-1} Q_{qp} V_{pp}) - i \operatorname{tr}(T_{qq} Q_{qq}^{-1} Q_{qp} V_{pp} Q_{pq} Q_{qq}^{-1}) + O(t^2),$$

and consequently

$$f_2 = e^{-\operatorname{tr}(U_{pp} V_{pp})} \propto 1 - 2i \operatorname{tr}(T_{pq} Q_{qq}^{-1} Q_{qp} V_{pp}) + i \operatorname{tr}(T_{qq} Q_{qq}^{-1} Q_{qp} V_{pp} Q_{pq} Q_{qq}^{-1}) + O(t^2). \quad \text{......(24)}$$

For the term $\frac{1}{2} nc$ merely contributes a numerical factor, and generally we have

$$e^{\operatorname{tr} X + O(x^2)} = 1 + \operatorname{tr} X + O(x^2).$$

On combining the results (23) and (24) we find

$$g^*(T) = 1 - 2i \operatorname{tr}(T_{pq} Q_{qq}^{-1} Q_{qp} V_{pp}) + i \operatorname{tr}\left[T_{qq}\left(Q_{qq}^{-1} Q_{qp} V_{pp} Q_{pq} Q_{qq}^{-1} + \frac{n-1}{n} Q_{qq}^{-1}\right)\right] + O(t^2). \quad \text{......(25)}$$

The two members of (25) are exactly equal, and not only proportional, since the first term on the right-hand side is equal to unity (§ 2). By comparing (25) with (18), p. 301, we can now read off the required mean values of the arrays in the Z-distribution, namely,

$$\left.\begin{aligned} V_{qp}^* &= -Q_{qq}^{-1} Q_{qp} V_{pp}, \\ V_{qq}^* &= Q_{qq}^{-1} Q_{qp} V_{pp} Q_{pq} Q_{qq}^{-1} + \frac{n-1}{n} Q_{qq}^{-1} \end{aligned}\right\}. \quad \text{......(26)}$$

The first of these relations becomes after transposition

$$V_{pq}^* = -V_{pp} Q_{pq} Q_{qq}^{-1}. \quad \text{......(27)}$$

It only remains to express the result in terms of R instead of Q. This can be done with the aid of (20) and (22), pp. 301, 302. Thus we finally get

$$V_{pq}^* = V_{pp} R_{pp}^{-1} R_{pq}, \quad \text{......(28)}$$

$$V_{qq}^* = R_{qp} R_{pp}^{-1} V_{pp} R_{pp}^{-1} R_{pq} + \frac{n-1}{n} (R_{qq} - R_{qp} R_{pp}^{-1} R_{pq}). \quad \text{......(29)}$$

It is remarkable that the value of V^*_{pq} should be independent of n, i.e. in the array $Z_{pp} = V_{pp}$ of the Z-population, the mean value of Z_{pq} is independent of the size of the sample (provided, however, that the inequality referred to on p. 298 is satisfied).

When $n \to \infty$, the matrices V^*_{pq} and V^*_{qq} become identical respectively with V_{pq} and V_{qq} which occur in the solution to Pearson's problem of selection (see equation (3), p. 295). In fact, we have even for finite n

$$V^*_{pq} = V_{pq},$$

and
$$\lim_{n\to\infty} V^*_{qq} = R_{qp}R^{-1}_{pp}V_{pp}R^{-1}_{pp}R_{pq} + R_{qq} - R_{qp}R^{-1}_{pp}R_{pq}$$
$$= R_{qq} - R_{qp}(R^{-1}_{pp} - R^{-1}_{pp}V_{pp}R^{-1}_{pp})R_{pq} = V_{qq}.$$

This proves Prof. Godfrey Thomson's conjecture regarding the connexion between statistical selection and arrays of samples.

REFERENCES

Aitken, A. C. (1935). "Note on selection from a multivariate normal population." *Proc. Edinb. Math. Soc.* (2), **4**, 106.

—— (1936). "A further note on multivariate selection." *Proc. Edinb. Math. Soc.* (2), **5**, 37.

Ingham, A. E. (1933). "An integral which occurs in statistics." *Proc. Camb. Phil. Soc.* **29**, 271.

Pearson, K. (1903). "On the influence of natural selection on the variability and correlation of organs." *Philos. Trans.* A, **200**, 1.

Thomson, G. H. (1939). *The Factorial Analysis of Human Ability.* Chapters xi and xii.

Wishart, J. & Bartlett, M. S. (1933). "The generalized product moment distribution in a normal system." *Proc. Camb. Phil. Soc.* **29**, 260.

THE CRANIAL AND OTHER SKELETAL REMAINS OF TASMANIANS IN COLLECTIONS IN THE COMMONWEALTH OF AUSTRALIA

By J. WUNDERLY, D.D.Sc.

1. Introduction

A research grant from the University of Melbourne enabled me to commence, in 1930, an enquiry into the physical anthropology of the extinct Tasmanians. The physical remains of these people available for examination consist chiefly of crania. In the last twenty-five years, a fairly large number of skulls claimed to be of Tasmanian origin has been added to public and private museum collections. The total number of crania contained in collections in the Commonwealth of Australia (including Tasmania) to which my data refer is 114. As 135 years have passed since the beginning of European settlement in Tasmania, it does not appear likely that many more crania will be unearthed in the island, unless a special search is made for them. The data from the material at hand should, therefore, be recorded for future reference, in case the specimens themselves be lost or damaged.

The aim of my investigation was to examine the conflicting records, opinions and methods of various authors, and to record original observations, with the object of presenting a true picture of the craniology of the Tasmanian aborigines. The basis of the enquiry, the methods employed, and the instruments used are referred to in appropriate sections of this paper.

The work associated with the enquiry was done in the Anatomy School of the University of Melbourne, in the public museums of Melbourne, Adelaide, Hobart, and Launceston, in the Institute of Anatomy at Canberra, and in the residences which contain the privately owned collections.

The data in this article have been carefully gathered in the hope that they will be of use to those who are competent to deduce from them information of value to all who are scientifically interested in the extinct Tasmanian race. I leave the more elaborate forms of biometric analysis to others, as my special knowledge is in anatomy, and not in statistics.

Our knowledge of the origin and migration of the extinct Tasmanian aborigines seems to be no further advanced to-day than it was in 1914, when Sir William Turner completed his classical enquiry into their physical characteristics. This fact is the more remarkable because many related enquiries have been made in the meantime. Turner not only made a sound, systematic and practical investigation of the physical characteristics of these people, but he also examined

and reviewed impartially almost all the writings on the subject published in English, French, or German up to his time.

The total physical remains of the extinct Tasmanian race are very small in quantity. They consist chiefly of about two hundred crania, many being in a bad state of preservation, and of not more than a dozen skeletons, a few dozen odd limb bones, and some odd specimens of hair and dried hands. Efforts to increase our knowledge of the physical anthropology of these people, therefore, depend mainly on investigations of their osteological remains.

Since the publication of "The Non-metrical Morphological Characters of the Tasmanian Skull" (Wunderly & Wood Jones, 1933) fourteen crania claimed to be Tasmanian have been added to collections in the Commonwealth of Australia. The number of crania included in my "Tasman" series is now 114, all of which have been systematically examined by me at least twice. A close study has been made of all the specimens in this series, in order to obtain a concise record of their morphological and anatomical characteristics. Particular attention has been given to the question of whether all the crania are authentic remains of Tasmanian full-blood aborigines or not, and an attempt has been made to classify the specimens correctly according to racial origin and sex. Special attention has been devoted to an important discovery of crania and limb bones at Eaglehawk Neck in Tasmania.

The metrical and anatomical data have been compared with those recorded by several investigators who have worked in Australia since 1897. Cranial anatomical characteristics observed during the present enquiry are listed and correlated with those recorded by Turner (1884 and 1908) with a view to building up a definitive basis for the racial diagnosis of skulls of Tasmanian full-blood aborigines.

2. An examination of some articles written since 1897 on the physical anthropology of the Tasmanians

Such widely diverse opinions are expressed in these articles that it was found necessary to enquire into the basis of each. Many of the articles are well-known and their merits are so obvious, and have already received such favourable comment, that reference is confined chiefly to the defects, if any, that have been found in them, or in the work on which some were based. Some of these defects could not have been discovered except through an examination of the specimens themselves. Correlation of all that has been found for, and against, the various articles enables one to assess the degree of reliability of the results and the reasonableness of the opinions expressed in them.

The articles examined, given in the appended list of references, deal with crania claimed to be Tasmanian, and contained in collections in the Commonwealth of Australia. The numbers of specimens treated are given in the following table.

Year of publication	No. of skulls examined	Investigators
1898	18	Harper & Clarke
1909–14	52	Berry & Robertson
1912–14	—	Buchner
1910	—	Cross
1916	3	Ramsay Smith
1924	6	Wood Jones & Campbell
1928	6	Hrdlička
1929	—	Wood Jones
1933	100	Wunderly & Wood Jones. 91 of the skulls were examined by the former and 6 by the latter
1935	—	Wunderly
1938	114	Wunderly (present article)

It was observed that although many of the writers refer to the work of Huxley, Turner, Duckworth, Broca, Topinard and others, yet only a few show evidence in their methods, or their writing, that they understood and could apply the teaching of these physical anthropologists. One way in which all but Harper & Clarke (1898), Hrdlička (1928) and Wunderly (1935) have failed to do so is in not adopting a critical attitude towards the material to be examined, in order to distinguish authentic from unauthentic specimens.

The investigations on which the articles have been based will be referred to separately.

(*a*) *Harper & Clarke* (1898). These anthropologists examined eighteen crania labelled "Tasmanian", which were contained in the Tasmanian Museum at Hobart, this being the first systematic investigation of its kind made in the Commonwealth. Their article contains abundant proof that they understood the work, and applied the teaching of Turner and others Great praise is due to them for having made a preliminary critical survey of the material available to them, which resulted in three skulls being rejected as "improperly classed", and three others being classified as the remains of half-castes. The craniometrical measurements recorded by Harper & Clarke were made directly on the skulls themselves, and they defined clearly the anatomical points between which the measurements were made. During the present enquiry their measurements were checked on the crania on two separate occasions, and neither a fault in their methods nor an inaccuracy in their records has been found. Their report is considered to be worthy of premier position among the earlier articles on the Tasmanians written by investigators in Australia. In the present enquiry the classification of the specimens examined by Harper & Clarke is consistent with theirs in so far as it draws a line between those which are, and the others which are not, the remains of full-blood Tasmanian aborigines.

(*b*) *Berry & Robertson* (1909 *a*, *b*, *c*); *Berry et al.* (1910, 1914). The names of these investigators are associated with those of a team of workers which included

a professor of anatomy, a medical graduate, two research mathematicians, and two medical students. Their reports have received such wide publicity that reference will be made to some aspects of their work which are not generally known.

Morant (1927) has already referred to defects which he found in Berry & Robertson's work. In marked contrast to Harper & Clarke, Berry & Robertson did not adopt a critical attitude towards the authenticity of the material available for examination. Among the fifty-two crania which they examined and accepted as authentic were all that were rejected by Harper & Clarke, or classified by them as the remains of half-castes. Other specimens included in the fifty-two have also been classified as unauthentic in the present enquiry, or as unsuitable as sources from which to obtain reliable data. Some of the latter kind have already been referred to and illustrated by Wunderly (1935). They took no account of the diagnostic anatomical characteristics of the Tasmanian skull as outlined by Turner (1884, 1908). Had they been familiar with Turner's teaching, they could hardly have failed to recognize the unauthentic specimens. They did not give an explicit account of the basis on which they judged all the crania to be authentically Tasmanian, beyond expressing the opinion that "every one presents over 90 % of the features so characteristically found in the skull of the Tasmanian aboriginal".

Berry & Robertson's descriptions of their methods contain many inconsistencies. For example, they state in some places that their measurements were made on the skulls, but in others they mention that many were made on dioptographic drawings, which were regarded by them as satisfactory sources from which to obtain accurate measurements. They placed such a high value on dioptographic drawings that special reference to them seems to be justified.

The Berry & Robertson team made 211 (1909c), and the writer has made 245, dioptographic drawings of Tasmanian crania. The same individual Martin's dioptographic apparatus was used in both cases. The instrument was tested by me for accuracy, and it was found that, after the most careful adjustment, the error in the drawing could not be reduced below 2 %, and it was frequently as high as 4 % of the direct measurement. In many of their drawings there is no indication that certain parts had been lost through damage. In such instances many writers have assumed that Berry & Robertson's drawings represent the intact skull. It is unreasonable to expect fine accuracy in dioptographic drawings because, in addition to the errors introduced by mechanical defects in the apparatus, there are the added sources resulting from the inking by hand over a pencilled line, and the uncertainty due to the width of the line of the drawing. My conclusion is that dioptographic drawings may be regarded as useful general representations of the various *normae* of a skull, but that they are not reliable as sources of measurements.

The measurements recorded by Berry & Robertson were checked on the specimens by me on at least two separate occasions. While the majority were found to be correct, many errors were discovered, some as great as 10%. Although Berry & Robertson could have measured the cranial capacity in twenty-four of the fifty-two crania referred to by them, it has been observed that they have only recorded it for the same specimens as those for which Harper & Clarke had already recorded it, with the exception of one which had been lost in the time between the two enquiries. Furthermore, Berry & Robertson's figures are the same as those of Harper & Clarke. These facts give one the impression that perhaps Berry & Robertson overlooked an obligation to acknowledge the use of Harper & Clarke's figures.

(*c*) *Büchner* (1912 *a*, *b*). Buchner was one of the mathematicians associated with Berry & Robertson. He depended on the dioptographic drawings for his metrical data, and disclosed his opinion of the drawings when he stated that the "diagrams are therefore strictly accurate and correlative". In a special enquiry made by him into the degree of prognathism of the Tasmanian skull, he again depended on measurements made on the drawings. In many of the skulls the bone in the region of the prosthion had been lost through damage before Berry & Robertson made the drawings, but this loss is not indicated by them. Consequently, many of Büchner's basio-alveolar measurements are not reliable, a defect for which the anatomist rather than the mathematician should be held responsible.

It should be noted that Büchner measured the nasio-alveolar, and the basio-alveolar diameters from a common point—the alveolar point—whereas separate points are used in the specifications of the International Agreement. The distinction between the alveolar point and the prosthion has been overlooked by a majority of those who have gathered craniometric data in Australia.

(*d*) *W. Ramsay Smith* (1916). There is little to add to the comments on the work and writing of W. Ramsay Smith which have already been made by me (1935). He examined two skulls and a fragment contained in the collection of the Australian Museum, Sydney, and one which was at that time in a private collection, but has since been lost. His article does not contain any reference to the work of Turner, or any evidence of knowledge of Turner's work and writing. Without questioning the authenticity of the material submitted to him, he accepted it as authentic, although the classification of one skull was based on no better evidence than a label which had been attached to it in the Sydney Museum, and which was inscribed "Tasmanian from Hobart". Turner emphasized the folly of relying on such labels. Smith accepted this skull as that of a male Tasmanian aboriginal. It does not possess any of the characteristics defined by Turner as indicative of masculinity, and the facial part exhibits clearly marked European characters. In the present enquiry it has been classified as the remains of a female mixed-blood (Australian-European).

Ramsay Smith did not specify the methods he used in measuring the crania, or compare his results with those obtained by other investigators.

(*e*) *Wood Jones & Campbell* (1924). These enquirers examined and described six skulls contained in the collection of the South Australian Museum in Adelaide. They took measurements on the skulls in accordance with the specifications of the International Agreement, except that they took both the nasio-alveolar and basio-alveolar diameters to the alveolar point. Their introductory remarks indicate that they seem to have over-estimated the value and accuracy of dioptographic drawings. They accepted, without question, all specimens as being authentically Tasmanian. This fact is of particular interest in the light of Hrdlička's later description of some as of "Australian type", and others as of "type quite Australian". In the present enquiry the specimens referred to by Hrdlička as resembling the Australian skull have been classified as not the remains of full-blood Tasmanian aborigines.

Wood Jones & Campbell's article does not refer to the work and writing of Turner or of any other physical anthropologist of note who had made a special study of Tasmanian craniology. The value of their craniometrical data is depreciated by the inclusion of measurements which are merely visual estimates of the correct dimensions. Because errors were found in some of the measurements recorded by them, check measurements were made on two separate occasions during the present enquiry.

(*f*) *Hrdlička* (1928). Hrdlička examined six crania claimed to be Tasmanian when he was in Australia in 1925 These specimens were contained in the Adelaide and the Melbourne public museums. He also examined thirty-one Tasmanian crania contained in the museum of the Royal College of Surgeons, London. While in Australia he measured and described a large number of Australian crania. His reports show that his knowledge of the craniology of the Tasmanians and the Australians was sufficiently sound to enable him to distinguish the differences—in some instances very small—between the crania of these races.

It is greatly to his credit that, in the short time available to him, he recognized that four of the six specimens referred to were not Tasmanian in type, but were of Australian type. These four crania had been previously accepted uncritically by Wood Jones & Campbell as the authentic remains of Tasmanian full-blood aborigines. In the present work the four specimens have been classed as the remains of Tasmanian-Australian mixed-bloods. This classification was based on the Tasmanian and the Australian diagnostic anatomical characteristics, and it is confirmed by the reliable evidence of ethnological remains and historical and geographical records.

(*g*) *Wood Jones* (1929*b*). An excellent suggestion was made by Wood Jones that composite drawings of the hypothetical skull, occupying the position of mean in a series of crania, would be useful in conveying an idea of the general

form of each racial type of skull.* If separate composite drawings had been prepared for each sex, their value would have been more than twice as great as that of a single drawing. It is unfortunate that the composite drawings prepared by Wood Jones were based on data obtained from the work of Berry & Robertson, and not from new data.

Summarizing the examination of the articles referred to in the light of the results of the present enquiry, it has been found that:

(i) the critical attitude adopted by Harper & Clarke and by Hrdlička towards the material which they examined was an essential preliminary step towards the ultimate correct classification of the crania in the Tasman series, and,

(ii) the metrical data provided in their articles are more reliable than those contained in the others.

3. The authenticity of the skulls

When the present enquiry began it was found that a relatively large number of skulls regarded as Tasmanian in origin had been added to public and private collections in the Commonwealth of Australia since 1909, when Berry & Robertson announced that the number was then fifty-three. A preliminary survey of material available for examination revealed that some specimens, although labelled "Tasmanian", are remains of Europeans. These skulls were doubtless unearthed in Tasmania, but they do not exhibit any evidence of aboriginal origin. For this reason they have not been included in the Tasman series of crania.

Still other specimens do not exhibit even slight resemblance to the Tasmanian, or any Negroid, or Negrito type of skull, the difference in a few cases being conspicuous to a gross extent. Occupying intermediate positions between these specimens and the skulls of general Tasmanian type are some crania which possess characteristics suggesting origin from either (*a*) two aboriginal races, or (*b*) a European and an aboriginal race, or (*c*) a Mongolian (Chinese) and an aboriginal race.

Turner concluded that some skulls examined by him had been regarded as authentically Tasmanian merely because they were labelled "Tasmanian" by collectors or museum officials with no special knowledge of craniology. He considered that some such specimens were remains of half-castes, with Polynesian or European admixture.

In the present enquiry all crania claimed to be Tasmanian have been included in what is called the "Tasman" series, providing they exhibit any evidence, however small, of origin from an aboriginal race; but the Tasman series has been divided into several sections, which enable the specimens of the

[* The numerous type contours for racial series of crania which have been given in papers in *Biometrika* since 1911 are composite drawings of the "mean type".—Ed.]

same racial origin, whether full-blood or mixed-blood, to be grouped together. Particulars of the individual specimens are given in Appendix I.

Reliable historical records, narratives, and official documents contain abundant reference to mating between Tasmanian aborigines and either Australian aborigines or Europeans. According to West (1852) and others, some Australian aborigines were sent to Tasmania in 1820 and 1828, owing to an official plan for pacifying the Tasmanians; inter-marriage between Tasmanians and these Australians was common. One Australian native was responsible for many murders, a victim being a Polynesian. Many of the Australians are known to have died in Tasmania. There is also reliable proof that "blackbirders", whalers and sealers were responsible for transporting natives, particularly women, to Tasmania from other localities. Many of these women gave birth to half-caste children while in Tasmania. It is also possible that some racial admixture occurred in the eighteenth century, or earlier, as the result of visits of adventurous explorers who left no records of their voyages owing to illiteracy or shipwreck.

Numerous photographs of groups of Tasmanian natives are available in which it is easy to distinguish the facial differences between full-bloods and mixed-bloods, particularly those of Tasmanian-European origin. Official records show that the mixed-blood inhabitants of Tasmania resulted from mating between two or more of the following races: Tasmanian, Australian, European, Chinese, Indian, Japanese, Maori, Negro, Polynesian, Syrian, and others.

The individual histories of a number of the crania emphasize the need for a critical attitude towards their authenticity. Since 1897 several specimens have disappeared from collections, some in suspicious circumstances. It is known that "trafficking" in Tasmanian crania has occurred on a few occasions, and that an unauthentic was substituted for an authentic specimen at least once. Prolonged enquiry has revealed that some supposed Tasmanian specimens were unearthed on the Australian mainland, while others were gathered from still more distant localities. Some of the latter belonged to private collectors residing in Tasmania, after whose death the collections were presented or sold to museums, unaccompanied by written records, or were divided among surviving relatives. The supposition that some such specimens are of Tasmanian origin has therefore no sound foundation. Ethnological specimens, geographically associated with the unearthing of some crania in Tasmania, are racially and culturally referable to races other than the Tasmanian. A description of them will be contained in an article on the origin of the Tasmanian race, now in course of preparation.

The crania comprising the Tasman series have been gathered principally through organized search or casual finding. Some, however, have been acquired as gifts or as the result of purchase. In some instances the persons from whom they were obtained recorded where they had been found, while in others this

information was not given. Not a few of the crania had been in several collections before reaching their present locations, and in some cases the names of the owners of those former collections cannot be traced. Persistent enquiry has elicited evidence which shows that some specimens had not been found in Tasmania.

The Tasmanian diagnostic anatomical characteristics, which were very clearly defined by Turner, were used as a basis for the essential classifications. Whenever collateral evidence, in the form of reliable ethnological, historical or official data, was available, it was consulted. It is not claimed that the classification has been made without error, but much time and effort has been expended in an attempt to classify the specimens accurately on the chosen basis.

4. The anatomical diagnosis of Tasmanian crania

From among many publications which had, at first, been regarded as suitable, one book by Duckworth (1904) and four papers by Turner (1884, 1908, 1910, 1914) were finally selected as a basis for such criteria. Duckworth was relied on as a guide in principles. The many anatomical characteristics exhibited specifically in crania, authentically the remains of Tasmanian full-blood aborigines, were so minutely observed by Turner that his comprehensive description of them is considered the best ever published.

Based on the thirty-six listed characteristics defined by Turner (1908), a primary classification was made to separate the skulls of Tasmanian full-bloods from others. Twenty-three additional characteristics have been gathered during the present enquiry, and have been used to supplement those of Turner; these have proved helpful in diagnosing remains of mixed-bloods. Some of the characteristics refer only to male, and others only to female, skulls. Any skull exhibiting over 75% of such characteristics is classed as the remains of a Tasmanian full-blood aborigine.

It is interesting to note that in all skulls classed as remains of Tasmanian mixed-bloods, the cranial part is Tasmanoid in general form, while the facial part shows the foreign characteristics; furthermore, it is in these foreign facial features of the skull that one sees the clue to the identity of the admixing race, whether European or otherwise.

The differences in certain anatomical features found in skulls of Tasmanian full-bloods, Tasmanian-European mixed-bloods, and Australian full-bloods are listed below. The data for Tasmanian full-bloods are taken from Turner's papers, for the mixed-bloods from my own observations, and for Australian full-bloods from Turner and others. Photographs of typical full-blood Tasmanian skulls are reproduced in Plates I–V and of skulls presumed to be those of Tasmanian-Australian half-bloods in Plate VI.

List of diagnostic Tasmanian and related cranial characteristics

No. of characteristic	Tasmanian full-blood (taken from Turner)	Tasmanian-European mixed-blood (Wunderly)	Australian full-blood (from Turner and others)
Norma verticalis			
1	Elongated and dolichocephalic; some ovoid or pentagonal	Ovoid, or anterior ovoid and posterior somewhat pentagonal	Elongated and ovoid
2	Parietal eminences prominent	Some show little prominence	Practically no prominence
3	Behind eminences width rapidly decreases to occiput	Less rapid in some	Decrease in width is much more gradual
4	Frontal eminences distinct	More fullness in frontal area	Forehead recedes more abruptly
5	Male skulls show triangular area anterior to bregma	—	(Not recorded for a series)
6	Male skulls have shallow depression lateral to this triangle	—	(Ditto)
7	Frontal breadth small compared with maximum cranial	Frontal breadth greater than usual	Less difference between frontal and maximum breadths
8	Skulls keeled along sagittal suture: keel usually limited to anterior one-third of suture	Generally less keeling	Generally keel more prominent, and extending along whole length of suture
9	Middle or posterior one-third of sagittal suture usually depressed	Depression well-marked in some	Depression unusual
10	Parietal foramina small or obliterated	Small	Small
11	Supra-inial region large and rounded in female skulls, small in males	Female fairly large and rounded, male not large	Not so noticeable in male or female
12	Inion not large	Sometimes large	Sometimes large
Norma lateralis			
1	Forehead recedes in males and more nearly vertical in females	Forehead generally fuller	Forehead more receding
2	Glabella and supraciliary ridges prominent in males	Less prominent	More prominent
3	Nasion deeply depressed	Some have little or no depression	Depression usually deep
4	Between obelion and lambda the vault slopes gradually downwards	Slopes less gradually	Slopes more gradually
5	Supraciliary ridge and upper border of orbit project in front of lower border	Usually level or upper behind lower	Level or upper behind lower
6	Outer border of orbit far behind inner border	Level in some	More nearly level
Norma facialis			
1	Vault roof-shaped	Roof-shaped or more rounded	Roof-shape more acute
2	Absence of grooves above supra-orbital foramina	Grooves in some	Grooves common
3	Maxillo-nasal spine diminutive	Spine prominent in some	More prominent
4	Breadth of anterior nares usually greater than half height	Breadth less, aperture almost parallel-sided in some	Breadth compared with height less than in Tasmanian, lower border 'guttered'

List of diagnostic Tasmanian and related cranial characteristics (cont.)

No. of characteristic	*Tasmanian full-blood* (taken from Turner)	*Tasmanian-European* mixed-blood (Wunderly)	*Australian full-blood* (from Turner and others)
Norma facialis (*cont.*)			
5	Nasal margins rounded	Margins very sharp	Less rounded
6	Canine fossae distinct and in some very deep (deeper in female than male skulls, J. W.)	Usually shallow	Deeper than in Tasmanian
7	Orbits low but wide	Orbits high, more nearly circular and borders sharp	Orbits approach square in many, more varied in shape than in Tasmanian
8	Infra-orbital suture usually obliterated	(Not recorded)	(Not recorded)
9	Malar bones small	Malar bones large in some	Malar bones large
Norma basalis			
1	Palate wide, and shallow to moderate height; none high	High and narrow in many	Wider and larger than in Tasmanian, often very high
2	Some exhibit fourth molar teeth	Not seen in any	Seen less often than in Tasmanian
3	No instance of artificial extraction of incisor tooth	Not seen	Common
4	No malocclusion of teeth (in present work a few instances of impaction of mandibular third molar teeth seen, J. W.)	Many show malocclusion of teeth	Impaction of mandibular third molars in some
Norma occipitalis			
1	Many have wormian bones in lambdoid suture	Less common	Not as usual as in Tasmanian
2	Inion not large	Large in some	Large in greater number than in Tasmanian
3	Superior curved occipital line prominent in some and divided into upper and lower lines in others	—	—
4	Third occipital condyle was not seen in any specimen	Third condyle not seen	(Not recorded)
5	Two skulls have external pterygoid plate fused with spine of sphenoid and pierced with two pterygo-spinous foramina	—	—

Supplementary diagnostic characteristics (Wunderly)

General			
1	Surface of bone very smooth	Some rough	Rougher than Tasmanian
2	Areas of attachment of muscles only slightly uneven	More uneven	All more uneven
3	All borders and margins rounded	Borders of facial part sharp	Not so rounded as in Tasmanian
4	General characteristics exhibited most clearly in distinctly masculine skulls	—	—
5	Closer resemblance between juvenile skulls of either sex and adult female than between latter and adult male skulls	—	—

List of diagnostic Tasmanian and related cranial characteristics (cont.)

No. of characteristic	Tasmanian full-blood	Tasmanian-European mixed-blood	Australian full-blood
Norma verticalis			
1	Superior temporal lines do not approach as close to sagittal suture as in the Australian	Not so close as in Tasmanian	Approach very close in many, especially males
2	Depression in sagittal suture (Turner) diamond-shaped, longer diameter coinciding with suture	Seen in some	Seen in some
Norma lateralis			
1	Middle one-third of each side of coronal suture slightly complicated in some skulls	—	Not complicated so often as in Tasmanian
2	In some specimens nasal bones weakly aquiline and markedly convex medio-laterally, very narrow at constriction	More aquiline or flatter, not narrowly constricted, some parallel-sided	Not so narrowly constricted as Tasmanian, generally wider, sides of some nearly parallel
3	Height of mandible behind second molar usually much less than symphysial height	Heights more nearly equal	Height behind second molar larger compared with symphysial, both larger than in Tasmanian
4	Squamous temporal flat antero-posteriorly and from above below, whole temporal fossa flat	Slightly convex in some to full in others	Flat or slightly convex
Norma facialis			
1	Very slight inclination, if any, between upper and lower borders of orbits	More inclination	Some show great inclination
2	Fronto-nasal and fronto-maxillary sutures usually almost straight	Fronto-nasal elevated in many	Straight or elevated above nasion
Norma basalis			
1	Maxillary palatal torus common	Not so common	Fairly common
2	Zygomatic arches thin medio-laterally	Usually thicker	Much thicker and rougher
3	Teeth not very large	Teeth smaller and degenerate, very little wear	Teeth larger, greater wear than in Tasmanian
4	Remarkable approach to uniformity of form and size in corresponding teeth	Wide differences	Not so uniform as in Tasmanian
5	Morphological elements more distinctly outlined than in teeth of Australians, and still more than in those of Europeans	Elements indistinct	Not so distinct
6	Majority of teeth show greater number of those elements than are seen in Australians or Europeans	Fewer elements	Not so many as in Tasmanian
7	Closer resemblance between Tasmanians' permanent and deciduous teeth (judged by the usually accepted descriptions of the latter) than seen in Australians or Europeans	Less resemblance	Not so much alike as in Tasmanian
8	Form of upper dental arch U-shaped	Variable in form and often irregular in shape	Usually parabolic or similar to elongated horse-shoe
9	Teeth occupy regular positions in each arch	Irregular to very irregular in position	Regular
10	Dental caries not found in any skull of aboriginal who lived in natural state, seen in many skulls of those who lived in contact with civilization	Extensive caries in large majority	Rare in natural state, common in contact with civilization

5. SKELETAL REMAINS FOUND AT EAGLEHAWK NECK

A description of the finding of aboriginal adult and juvenile skeletal remains at Eaglehawk Neck, on the east coast of Tasmania, in 1919, was published by Lord (1919), from whose paper the following extract is taken.

> Upon arrival at Eaglehawk Neck, in company with Mr Brister and Mr W. H. Clemes, I found that a slight sandslip had occurred on the south-eastern face of one of the large sand dunes forming Eaglehawk Neck. A number of small bones appeared on the surface, and after collecting these a start was made to examine below the surface. Upon excavation a number of larger bones and several skulls were revealed. Owing to the fact that the dune in question was covered with Boobialla (*Myoporum insulare*), and the roots in many cases completely filled the cavities of the bones, the task of exhuming these relics of a bygone race was one of considerable difficulty.

The bone in all the specimens is extraordinarily clean, a condition no doubt due to their burial in sand. Unfortunately, a large majority of the bones were broken during the difficult exhumation, and, although the cranial part of some of the skulls is intact, the fragments of the facial parts cannot be identified as belonging to any particular cranium. Limb bones were also found, but many are broken. All these specimens are deposited in the Tasmanian Museum, Hobart; a list of 330 of them was published by Lord & Crowther (1920). Since there is no evidence of ante-mortem injury to indicate death by fighting, it is probable that a tribal group perished from some natural cause.

Five of the crania are sufficiently well preserved to enable reliable racial diagnoses to be made, and also to provide anatomical and metrical data of value. They are numbered as follows:

	Male	Female	Juvenile
Tasman series Nos.	79, 80, 81	78	82

In addition to the specimens included in the Tasman series, all fragments of facial and cranial parts were examined. The Tasmanian anatomical characteristics described by Turner are clearly exhibited in the Eaglehawk Neck remains, a few being more marked than in any other specimens. Not a single characteristic was found, whether facial or cranial, that would suggest either admixture, or racial origin other than Tasmanian. Turner noted that, while all skulls of Tasmanian full-bloods examined by him bear a general resemblance to one another, yet minor differences occur in individual specimens; for instance, the cranium viewed in *norma verticalis* may be elongated and dolichocephalic, ovoid or pentagonal. Similar differences seen in the Tasman series of skulls have been noted during the present enquiry in the case of cranial form, form of the orbit, the region of the forehead and nasion, and several other characteristics.

Because the cranial remains found at Eaglehawk Neck were found simultaneously and in the one locality, it was decided to compare them with the other crania classified as those of Tasmanian full-bloods in the Tasman series. When these crania are roughly divided into two groups, the one containing skulls of aborigines known to have died since European settlement began, and the other containing those unearthed in the earlier days of settlement, it is found that the Eaglehawk Neck specimens resemble the latter more closely than the former group. In these two groups ("old group" and "recent group" in the following table), the better-known Tasmanian characteristics differ to some extent, as shown below:

Characteristic	Old group	Recent group
Cranial size	Larger	Smaller
Cranial form	Curvilinearly pentagonal	Angularly pentagonal
Orbit	Somewhat rectangular	Markedly rectangular
Nasion	Depressed	Deeply depressed
Parietal eminences	Prominent, rounded	Prominent, angular

The general difference between the "old" and the "recent" group of skulls suggests that the latter exhibit greater specialization, due perhaps to long occupation in a restricted insular environment. The general difference between the two groups is not regarded as indicative of a difference in racial impurity.

The maximum lengths of the three Eaglehawk Neck skulls of males occupy the first, second and fourth places, respectively, in the table of measurements (Appendix III) of the crania of male Tasmanian full-bloods. The maximum breadths of their vaults occupy the first, fifth and tenth positions, respectively. The female specimen fills the second place among the skulls of the female Tasmanian full-bloods in the case of the maximum length, and, with three other female specimens, it shares the sixth place in the case of the maximum breadth of the vault. It is therefore apparent that the Eaglehawk Neck crania are within the recognized metrical limits of Tasmanian crania so far as size is concerned.

Turner pointed out that in many Tasmanian skulls a part of the sagittal suture lies in a depression between two lateral ridges. This characteristic is well marked in the Eaglehawk Neck skulls, the depth of the depression being 5 mm. in No. 80 of the Tasman series. Individual measurements of the femora and tibiae from Eaglehawk Neck are given in Appendix II.

6. The Tasman series of crania: metrical data

Particulars of the crania in this series—their present locations, the related individual reference numbers, and the related racial and sexual classifications—are given in Appendix I. Measurements of the specimens accepted as representing

full-blood Tasmanians and also those of Tasmanian-Australian half-bloods are given in Appendix III

The basis on which the classification has been made in the present enquiry has already been mentioned. It is believed that the classification is generally reliable to the extent that it separates the crania of Tasmanian full-blood aborigines from those of other origin, whether full-blood or mixed-blood. As regards two specimens, which have been included with the skulls of the Tasmanian full-bloods, there is a small doubt, though the evidence is not considered sufficient to justify their exclusion. The recognition of crania of Australian full-bloods, and a majority of those classified as the remains of Tasmanian-European mixed-bloods, has presented no difficulty, but some uncertainty exists as to whether some of the latter skulls are remains of Tasmanian-European or Tasmanian-Chinese mixed-bloods.

Seven out of eight mandibles unassociated with crania have been classified as remains of Tasmanian full-bloods, because there is not sufficient evidence to exclude them. One mandible has been classified as that of an Australian full-blood; its rugged construction, large and greatly worn teeth, and the form of its dental arch all differ from the corresponding features in authentic Tasmanian mandibles.

Thanks to Turner's descriptions, most of the skulls have been easy to classify according to sex. A small minority proved difficult, but it was considered preferable to attempt to classify correctly each skull, rather than to relegate any to a group of specimens of unassigned sex. One or two regarded as the remains of female Tasmanian full-blood aborigines may be those of males: their anatomical characteristics indicate femininity, while their cranial capacity suggests masculinity.

Of the 114 specimens in the Tasman series—all of which are in Commonwealth collections, it should be remembered—I took measurements of 101, the remaining thirteen being too fragmentary for the purpose. The individual readings for the fifty-eight adult skulls judged to be those of full-blood Tasmanians, and the eight adult skulls judged to be those of Tasmanian-Australian half-bloods are given in Appendix III. Of the total series of 114 specimens, Berry & Robertson have published measurements of fifty-two, Harper & Clarke of fifteen, Hrdlička of seven, Wood Jones & Campbell of six and Ramsay Smith of three.

Means derived from my measurements of the full-blood Tasmanian skulls in the Tasman series are compared in Table I with means given by Morant (1927) which were obtained by pooling the measurements provided by a number of earlier investigators.* It should be realized that the latter set is partly based

[* Comparisons between the two sets of means are made in the Note by Dr Morant appended to Dr Wunderly's paper.—Ed.]

on data for a few specimens which are not classed as full-blood Tasmanian by me, and also on a considerable number in European collections.

Comparisons are made in Table II between a few means derived from my measurements and those given by other workers. The values given earlier relate partly to specimens in Commonwealth collections which I do not accept as full-blood Tasmanian (Harper & Clarke's and Hrdlička's) and partly to specimens in European collections (all Turner's and most of Hrdlička's). Most of the numbers are far too small to provide reliable means, but, nevertheless, a remarkably close agreement is found. It may be noted that the means for the very short Tasmanian-Australian mixed-blood and for the Australian full-blood series fall on the same side of all the other means in the case of the male and female cephalic and height-length and of the male nasal index.

The non-metrical morphological characters of Tasmanian crania

These characteristics, for material available in 1933, were recorded by Wunderly & Wood Jones (1933). Owing to the discovery of additional specimens, a revision of the data has been found necessary. In the present paper the characteristics are recorded only for the skulls of Tasmanian full-bloods (Tasman series, Section A). To make them more useful for purposes of reference, they have been recorded for each sex separately. The particulars in the former report, which are still applicable to all specimens now available, are not included in the present account. The directions given by Wood Jones (1929*a*) were again followed when recording the revised data.

(i) *Cranial form (fifty-seven crania of Tasmanian full-bloods)*

Reference has already been made to the cranial type of the Tasmanian skull, and to the two modifications in this type which have been observed in the present work,

Norma verticalis. (*a*) The specialized form of skull, as seen particularly in the remains of the aborigines who died since the time of European discovery, is generally pentagonal, with "pronounced bosses situated far posteriorly on the parietal bones, and a relatively small minimum frontal breadth. The occipital region is broad, and well rounded, but in some specimens it is small in area and prominent. The medio-lateral thickness of the zygomatic arch is remarkably small compared with that of the Australian".

(*b*) In the Eaglehawk Neck skulls and some others which resemble them fairly closely it is seen that the general outline form is not so markedly pentagonal, the parietal eminences are not so acutely prominent, and they are not situated so far posteriorly. In short these skulls are more gently rounded, and they do not exhibit the features which may be termed "outline angularities" that distinguish the specialized Tasmanian skull.

TABLE I

*Mean measurements of series of Tasmanian skulls**

	Male		Female	
	Tasman series (Wunderly)	Pooled series (Morant)	Tasman series (Wunderly)	Pooled series (Morant)
Max. glabella occipital length (L: 1)	185·4 (30)	182·2 (43)	177·9 (25)	174·6 (20)
Glabella-inion length (2)	180·0 (30)	177·7 (36)	172·3 (25)	166·3 (16)
Maximum breadth (B: 3)	138·2 (27)	136·0 (60)	135·8 (25)	132·4 (36)
Max. frontal breadth (B'': 6)	111·0 (25)	108·2 (24)	108·4 (25)	103·6 (10)
Max. bimastoid breadth (7)	119·9 (26)	—	116·7 (22)	—
Min. frontal breadth (B': 5)	94·7 (25)	94·0 (62)	92·9 (27)	90·1 (35)
Basio-bregmatic height (H': 4a)	129·8 (24)	130·9 (55)	129·2 (22)	125·3 (35)
Auriculo-bregmatic height (βOH: 4b)	114·1 (25)	—	111·6 (23)	—
Chord nasion-basion (LB: 9)	98·1 (22)	98·8 (55)	94·6 (22)	92·7 (34)
Chord prosthion-basion (10)	100·3 (9)	—	97·1 (10)	—
Length of foramen magnum (fml: 21a)	35·6 (22)	35·7 (53)	34·4 (22)	34·2 (31)
Breadth of foramen magnum (fmb: 21b)	29·0 (22)	29·6 (44)	29·0 (23)	28·4 (27)
Horizontal circumference (U: 23a)	515·8 (25)	511·3 (48)	500·1 (23)	489·5 (23)
Arc nasion-bregma (S_1: 22 (i))	128·8 (27)	127·2 (44)	125·1 (26)	121·3 (23)
Arc bregma-lambda (S_2: 22 (ii))	131·0 (28)	126·2 (42)	127·6 (24)	122·2 (23)
Arc lambda-opisthion (S_3: 22 (iii))	112·9 (23)	111·8 (37)	110·5 (22)	109·4 (16)
Arc nasion-opisthion (S: 22)	370·3 (21)	365·8 (36)	364·3 (23)	350·5 (15)
Broca's transverse arc (23)	293·2 (25)	290·2 (40)	286·7 (23)	283·5 (17)
Chord nasion-alveolar point ($G'H$: 12)	62·4 (12)	62·5 (36)	61·1 (11)	59·9 (16)
Orbito-alveolar height (20)	38·8 (19)	—	36·7 (16)	—
Bizygomatic breadth (J: 8)	130·4 (9)	131·0 (44)	126·6 (12)	122·0 (21)
Flower's interorbital breadth (15)	22·7 (25)	25·3 (20)	22·0 (21)	23·8 (13)
Dacryal orbital breadth, R ($O_1'R$: 16)	38·1 (19)	39·3 (40)	36·9 (20)	38·3 (18)
Dacryal orbital breadth, L ($O_1'L$: 16)	37·7 (19)		37·0 (18)	
Orbital height, R (O_2R. 17)	29·9 (19)	31·05 (60)	30·9 (20)	31·7 (31)
Orbital height, L (O_2L: 17)	29·3 (19)		30·6 (20)	
Nasal height (NH: 13)	45·1 (21)	47·1 (58)	44·7 (19)	44·9 (30)
Nasal breadth (NB: 14)	26·9 (19)	27·8 (57)	25·9 (20)	26·3 (29)
Width of alveolar border (18)	66·0 (14)	—	63·9 (14)	—
Height of alveolar curve (18a)	60·7 (9)	—	57·1 (11)	—
Breadth of palate (G_2: ?19b)	39·5 (15)	—	37·6 (13)	—
Length of palate (G_1': 19a)	49·6 (10)	—	48·7 (11)	—
Minimum thickness	4·4 (25)	—	3·9 (22)	—
Maximum thickness	7·5 (25)	—	6·7 (22)	—
Capacity (C: 24)	1247·1 (14)	1264·3 (33)	1242·8 (14)	1153·8 (25)
100 B/L	74·2 (27)	74·2 (43)	76·4 (24)	75·1 (19)
100 H'/L	70·6 (24)	71·3 (37)	72·3 (22)	71·1 (19)
100 B/H'	105·8 (22)	103·9 (55)	106·1 (21)	105·7 (34)
100 B'/B	69·0 (24)	—	68·7 (25)	—
100 O_2/O_1', R	78·5 (19)	79·4 (40)	83·8 (20)	83·3 (17)
100 O_2/O_1', L	77·8 (19)		82·5 (18)	
100 NB/NH	59·9 (19)	59·1 (57)	58·6 (19)	59·0 (29)
100 fmb/fml	81·6 (22)	82·1 (42)	84·7 (22)	83·3 (26)

* Measurements for which both male and female means of the Tasman series are based on fewer than ten skulls are omitted. See p. 335 for remarks on the definitions of the measurements.

TABLE II

Mean measurements for Tasmanian and Australian series of skulls

		Tasmanian full-blood (Wunderly)	Tasmanian (Harper & Clarke)	Tasmanian (Turner)
♂	Cephalic index (100 B/L)	74·2 (27)	74·0 (6)	72·5 (8)
	Height index (100 H'/L)	70·6 (24)	70·0 (4)	72·0 (7)
	Orbital index (100 O_2/O_1')	78·1 (19)	79·4 (6)	77·3 (7)
	Nasal index (100 NB/NH)	59·9 (19)	54·0 (6)	59·8 (7)
	Capacity	1247 (14)	1282 (3)	1235 (7)

		Tasmanian (Hrdlička)	Tasmanian-Australian mixed-blood (Wunderly)	Australian full-blood in Tasman series (Wunderly)
♂	Cephalic index (100 B/L)	74·1 (22)	71·7 (5)	70·4 (3)
	Height index (100 H'/L)	—	69·4 (5)	69·7 (4)
	Orbital index (100 O_2/O_1')	80·3 (21)	79·6 (5)	74·7 (4)
	Nasal index (100 NB/NH)	56·7 (20)	51·8 (4)	52·1 (4)
	Capacity	—	1285 (3)	1261 (2)

		Tasmanian full-blood (Wunderly)	Tasmanian (Harper & Clarke)	Tasmanian (Turner)
♀	Cephalic index (100 B/L)	76·4 (24)	77·0 (5)	74·2 (1)
	Height index (100 H'/L)	72·3 (22)	72·5 (4)	73·0 (1)
	Orbital index (100 O_2/O_1')	83·1 (18)	84·8 (4)	84 6 (1)
	Nasal index (100 NB/NH)	58·6 (19)	55·2 (3)	61·0 (1)
	Capacity	1243 (14)	1089 (5)	1260 (1)

		Tasmanian (Hrdlička)	Tasmanian-Australian mixed-blood (Wunderly)	Australian full-blood in Tasman series (Wunderly)
♀	Cephalic index (100 B/L)	76·2 (15)	73·1 (2)	73·3 (6)
	Height index (100 H'/L)	—	68·8 (2)	70·9 (5)
	Orbital index (100 O_2/O_1')	84·2 (15)	84·3 (3)	85·9 (6)
	Nasal index (100 NB/NH)	58·4 (15)	55·7 (3)	56·9 (6)
	Capacity	—	1172 (2)	1077 (5)

Norma lateralis. The Eaglehawk Neck specimens, viewed from this aspect, are seen to be more rounded than the specialized skulls. The temporal fossae in the Eaglehawk Neck skulls are usually a little fuller than in the specialized specimens, in some of which they are notably flat.

Norma facialis. The facial margins of the orbit in the Eaglehawk Neck crania do not form such a pronounced rectangle as is seen in many of the skulls of specialized form.

Norma occipitalis. The angularities seen in the irregular pentagonal outline of the specialized skull are not so noticeable in the Eaglehawk Neck specimens, although the general outline seen in the one form closely resembles in other respects that seen in the other form. The "depression" of the posterior one-third or one-half of the sagittal suture is deeper in the Eaglehawk Neck male specimens than in any other skulls in the Tasman series.

(ii) *Cranial asymmetry*

It is now possible to demonstrate the asymmetry of a skull graphically, and in a rough quantitative way, by means of a modified Schwarz drawing apparatus.

(iii) *Sutures (fifty-seven crania)*

The only alteration necessary with regard to this characteristic is in respect of the total number of crania for which the particulars are now applicable.

(iv) *Ossa suturarum (fifty-two crania)*

	Males 28 skulls	Females 24 skulls
Total number of ossicles	64	77
Average per skull	2·3	3·2
Skulls having ossicles:		
Bi-laterally in lambdoid suture	28%	37%
Unilaterally in lambdoid suture	43%	29%
In occipito-mastoid suture	21%	16%
At asterion	11%	33%
Percentage with ossicles in the		
Lambdoid suture	84	67
Right lambdoid	36	36
Left lambdoid	36	30
At lambda	12	1

One female skull has an ossicle in the right half of the coronal suture; one female has one in the sagittal suture, and another has four in the same suture. The largest number of ossicles observed in any skull is thirteen in a female specimen.

(v) *Pterion (forty-eight crania)*

The pterion is of normal contact bilaterally and of usual size in 28 % of the twenty-five male skulls, and in 22 % of the twenty-three female skulls; and wide in 8 % of the male skulls, and in 4 % of the female skulls. The contact in 4 % of the male skulls and in 9 % of the female skulls is seen to be normal on each side, but of the usual size on one side and narrow on the other. Epipteric bones completely occupy the pterion bilaterally in 12 % of the male and in 4 % of the female skulls. Two female crania exhibit the pithecoid contact on each side. An epipteric bone unilaterally accompanied by a normal contact of usual size appears in 20 % of the male and in 30 % of the female skulls. In one female skull the normal contact of usual size is associated with a normal, wide contact on the other side, and in another female specimen the contacts on each side are fused.

The pterion of the side remaining in skulls in which the parts of the other side are lost through damage is found to be as follows:

(*a*) normal and of usual size in four male and one female skull,

(*b*) normal and narrow in one male specimen,

(*c*) by epipteric bone in one male skull.

(vi) *Epipteric bones (forty-eight crania)*

These bones were found bilaterally in 12 % of the twenty-five male skulls and in 4 % of the twenty-three female skulls. They were observed unilaterally on the left in 24 % of the male and in 30 % of the female skulls. One female specimen has an epipteric on the right only, but this condition was not seen in any male skull.

(vii) *Supra-orbital foramina, notches or grooves (fifty-eight crania)*

Bilateral grooves, some being shallow, were found in 39 % of the thirty-one male skulls; a foramen on one side and a notch on the other in 13 %; a groove with a small accessory foramen bilaterally in 13 %; a groove on one side and a notch on the other in 10 %; and a notch and an accessory foramen bilaterally in 6 %. One skull has a notch bilaterally, while another has a groove on one side and on the other a groove with an accessory foramen. In four skulls in which the parts on one side are missing, the other exhibits a notch in three specimens and a groove in one.

In the twenty-seven female skulls 30 % have a groove bilaterally, and 30 % a notch bilaterally. Each of the following conditions was observed in each specimen in four different groups consisting of two female skulls each:

(*a*) a foramen on one side and a notch on the other,

(*b*) a notch and an accessory foramen bilaterally,

(c) a groove and an accessory foramen on one side and a notch and an accessory foramen on the other,

(d) a notch on one side and a notch and an accessory foramen on the other

(viii) *Anterior ethmoid canal (thirty-four crania)*

In the male skulls it was found bilaterally in the suture in 58 % of cases, and in the frontal bone and independent of the suture in 10 %. In five male specimens in which the parts of one side are lost through damage, the canal is in the suture of the other side. In one male skull it was seen bilaterally in the frontal bone and confluent with the suture.

In 80 % of the female skulls it is situated in the suture bilaterally, and in one specimen only it is in the frontal bone bilaterally and independent of the suture. In one female specimen it was found in the frontal bone on one side and in the suture on the other. In one specimen in which the parts of one side had been lost the canal on the other side is in the suture.

(ix) *Sutures of the inner wall of the orbit*

An abnormal arrangement of these sutures was not observed in any skull classified as that of a Tasmanian full-blood.

(x) *Spheno-maxillary fissure (thirty-eight crania)*

This fissure was classified as narrow in 37 % of the male and in 21 % of the female skulls; as of moderate width in 53 % of the male and 63 % of the female specimens, and as wide in 10 % of the male and 16 % of the female crania.

(xi) *Form of the orbit*

The only particulars regarding the Tasmanian orbit which need be added to those already published are those concerning the orbit of the Eaglehawk Neck group of crania. The markedly rectangular form of the orbit applies to the specialized Tasmanian natives, but its form in the crania of the aborigines who are believed to have been some of the earliest inhabitants of the island was less noticeably rectangular.

(xii) *Infra-orbital foramen (forty-four crania)*

It was decided to classify the foramina separately from the independent sutures. A single foramen of usual size was found bilaterally in 50 % of the twenty-two male and also in 50 % of the twenty-two female skulls. A single foramen on one side, and a single foramen accompanied by a small accessory foramen on the

other side, was found in 36 % of the male, and in 14 % of the female crania. The following conditions were found in the numbers of specimens indicated:

	Males	Females
(*a*) Single and an accessory foramen bilaterally	1	1
(*b*) Parts on one side lost, remaining side shows single normal foramen	1	4
(*c*) Double foramen bilaterally	1	—
(*d*) A single normal foramen and two accessory foramina on one side, and a single foramen and one accessory foramen on the other	—	1
(*e*) A single normal foramen and two accessory foramina on one side, and a single normal foramen on the other	—	1
(*f*) Parts on one side lost, the remaining side exhibits a single normal foramen and one accessory foramen	—	1

A complete independent suture from a foramen to the orbital border was found bilaterally in three male and nine female crania. In four female skulls it is present unilaterally, and a complete suture on one side and an incomplete suture on the other was observed in one female specimen.

(xiii) *Form of the jugal*

The remarks already published still hold good.

(xiv) *The nasal bones (forty-seven crania)*

The nasal bones were found to be normal and symmetrical in 36 % of the twenty-five male crania, and in 32 % of the twenty-two female specimens; normal and asymmetrical in 20 % of the males and 23 % of the females; narrow and symmetrical in 28 % of the males and 14 % of the females; narrow and asymmetrical in 4 % of the males and 9 % of the females; wide and symmetrical in 8 % of the males and 18 % of the females; and wide and asymmetrical in 4 % of the males and 4 % of the females. The internasal suture is fused in one and partly fused in three male skulls.

(xv) *The narial aperture (thirty-nine crania)*

The specimens in which the lateral margins of the aperture are "almost parallel-sided" were found to be unauthentic.

(xvi) *The nasal septum*

This is present in only one male skull, in which it is slightly deflected, and in three female crania, in two of which it is normal and in the other deflected.

(xvii) *The foramen ovale* (*fifty crania*)

The average size of this foramen was found to be from 5 to 6 mm. long, by 3 mm. wide. The following conditions were found in the twenty-seven male and twenty-three female skulls:

	Males	Females
(*a*) Foramen complete and of average size	59%	39%
	Specimens	
(*b*) Complete and small bilaterally	3	2
(*c*) Incomplete and confluent with the *foramen spinosum* on one side, and complete and of average size on the other	1	2
(*d*) Incomplete and confluent with the *foramen spinosum* bilaterally	3	—
(*e*) Incomplete on one side and complete and of average size on the other	2	1
(*f*) The parts on one side are lost, and on the other the foramen is complete and of average size	2	2
(*g*) Ditto and the foramen on the remaining side is incomplete	1	—

The following conditions were observed once in female skulls:

(*a*) the parts on one side are lost and on the remaining side the foramen is incomplete and confluent with the *foramen spinosum*,

(*b*) ditto and on the remaining side the *foramen ovale* is complete and round,

(*c*) complete and round bilaterally,

(*d*) complete and round unilaterally, and on the other side the foramen is incomplete and confluent with the *foramen spinosum*.

(xviii) *The foramen of Vesalius* (*thirty-seven crania*)

In the seventeen male skulls the foramen is present and complete bilaterally in 41 %; absent bilaterally in 12 %, and present and complete unilaterally in 47 %. In the twenty female crania it is present and complete bilaterally in 50 %; absent bilaterally in 25 %, and present and complete unilaterally in 20 %. In one specimen it is present and incomplete bilaterally.

(xix) *Foramen spinosum* (*forty-nine crania*)

Only the external orifice of the *canalis spinosus* was examined. In the twenty-six male skulls it is complete bilaterally in 35 %, and unilaterally in 31 %, while in 15 % it is incomplete bilaterally. In four skulls in which one side has been lost the foramen on the remaining side is complete in three and incomplete in one. This incomplete foramen is confluent with the *foramen ovale*,

and one of the complete foramina has a double orifice. The double orifice is seen in two male skulls, and in four the confluence between the *foramen spinosum* and the *foramen ovale* is noticeable.

In the twenty-three female crania it is seen to be complete bilaterally in 30 %, and unilaterally in 26 %, while in 22 % it is incomplete bilaterally. In five skulls in which the parts of one side are lost the foramen on the remaining side is complete in three and incomplete in two. In three specimens the incomplete foramina are confluent with the *foramen ovale*, while in another three the complete foramina are situated high on the *spina angularis*. One skull exhibits a double orifice bilaterally.

(xx) *Spina angularis sphenoide* (*fifty crania*)

In the twenty-seven male skulls it is short and blunt bilaterally in 44 %, and short and blunt one side and short and sharp on the other in 18 %. Each of the following conditions was observed twice in the male skulls:

(*a*) short and sharp bilaterally,

(*b*) conical bilaterally.

One specimen exhibits a blunt arrow-head bilaterally and another skull possesses a long sharp arrow-head bilaterally. In four skulls in which the parts on one side are lost the *spina* on the other side is short and blunt in two, short and sharp in one, and conical and sharp in one.

In the twenty-three female crania it is short and blunt bilaterally in 43 %, and short and sharp bilaterally in 13 %. In two specimens it is short and blunt on the one side, and short and sharp on the other. One specimen has a broad and flat *spina* bilaterally and another a short bifid *spina* bilaterally.

In each of five skulls in which the parts of one side are lost the other side exhibits the following conditions:

(*a*) short and sharp *spina*,

(*b*) the *spina* consists of a high ridge,

(*c*) short and blunt,

(*d*) sharp arrow-head,

(*e*) long and sharp.

In one specimen a short sharp arrow-head is seen on one side and a sharp bifid *spina* on the other.

(xxi) *Laminae pterygoidei* (*forty-four crania*)

The attached margin of the lateral laminae fades away as a ridge close to the anterior margin of the *foramen ovale* in 64 % of the twenty-five male skulls. In 12 % it fades laterally, and in 8 % medially to the foramen. It is medial to the foramen bilaterally in one skull. In each of three crania in which the parts on one side are lost, the other side shows the ridge ending at the anterior margin of the foramen.

In 74 % of the nineteen female crania the ridge ends anterior to the foramen. Each of the following conditions was observed once:

(*a*) the ridge is lateral to the foramen bilaterally,

(*b*) the ridge is medial to the foramen bilaterally,

(*c*) the ridge is medial on one side and lateral on the other,

(*d*) the ridge ends at the anterior margin of *foramen ovale* on one side, and is lateral to the foramen on the other,

(*e*) the parts on one side are lost, and on the other side the ridge ends anterior to the foramen.

(xxii) *The jugular fossa and foramen* (*thirty-nine crania*)

In 62 % of the twenty-one male skulls the foramen on the right is larger than that on the left; in 14 % they are equal in size, and in 24 % the left foramen is larger than the right. In the eighteen female crania the right foramen is larger than the left in 72 %; they are equal in size in 17 %, and in 11 % the left foramen is larger than the right.

(xxiii) *The tympanic region* (*forty-eight crania*)

In 63 % of the twenty-seven male skulls the conditions found were classified as normal bilaterally. Exostoses in the external auditory meatus were observed bilaterally in 11 %. The floor of the mouth of the external auditory meatus was regarded as thick bilaterally in two skulls, and thin bilaterally in two others. The following conditions were observed once in different specimens:

(*a*) the margins of the meatus are rough bilaterally,

(*b*) a rough ridge occurs on the lower surface bilaterally,

(*c*) the parts of one side are lost and the other side exhibits a ridge higher than the *spina angularis*.

Greater variability was seen in the female crania. In 29 % of the twenty-one female specimens the conditions were classified as normal bilaterally and in 29 % the margins of the meatus are rougher than usual bilaterally. The bone of the floor of the meatus is thick bilaterally in 14 %, and thin bilaterally in 14 %. In one skull exostoses in the external meatus bilaterally were observed. One cranium shows exostoses unilaterally, and on the other side the conditions are normal. A rough ridge on the lower surface is exhibited bilaterally in one skull.

(xxiv) *Foramen of Huschke*

Not observed in any skull.

(xxv) *Styloid process* (*twenty-five crania*)

In the thirteen male crania it is short and small in cross-section bilaterally in 69 % of cases, and rudimentary in two specimens. It was recorded as being short and thick bilaterally in one skull. In one specimen in which the parts on one side are lost the process on the remaining side is rudimentary.

In 66 % of the twelve female skulls it is short and small in cross-section bilaterally, and in two specimens it is rudimentary bilaterally. In one skull it is short and thick bilaterally, and short and small in section unilaterally in one specimen in which the parts on one side are lost.

(xxvi) *The posterior condyloid foramen* (*forty-one crania*)

In the twenty male crania it is present unilaterally, on the right side, in 40 % of cases; bilaterally in 20 %; unilaterally, on the left side, in 5 %; and absent from both sides in 35 %. It is present bilaterally in 29 % of the twenty-one female skulls, present unilaterally in 42 %, and absent bilaterally in 29 %.

8. Summary

The author has investigated, during the past eight years, the physical characteristics of the extinct Tasmanian aborigines. The present article is a report on the anatomical aspects of the enquiry.

The remains of the Tasmanians consist chiefly of crania and a small number of other bones. A critical anatomical examination of the crania claimed to be of Tasmanian origin and contained in collections in the Commonwealth of Australia, together with reliable collateral evidence, reveals that some of the specimens are not authentic. The basis of the anatomical diagnosis of their racial origin is described.

A critical examination is made of several reports published since 1898 and their values are assessed.

The crania examined are numbered in a series known as the Tasman series, their numbers being related to those allotted in other enquiries. The classification of racial origin has been tabulated (Appendix I) to show its relation to that adopted by each of several other enquirers.

Special reference is made to osteological remains found at Eaglehawk Neck, and the probable significance of their particular characteristics is discussed.

The non-metrical morphological characteristics of the crania classified as those of Tasmanian full-bloods are recorded.

Turner gave such a comprehensive account of Tasmanian craniology that little can be added to it as the result of the present enquiry. He inferred that the Tasmanians were direct descendants from a primitive Negrito stock which had migrated across Australia. He also considered that they had become specialized in many ways as the result of long isolation. The observed intra-racial differences between the Eaglehawk Neck, or "old", type of crania and the "recent" type may indicate progressive specialization, resulting from long occupation in a restricted environment. These differences constitute the only anatomical evidence found, during this enquiry, which has a bearing on the length of time during which the Tasmanians inhabited their island. The extent of the differences seems to point to the probability of a lengthy time period.

During the present enquiry no instance of cranial deformity of any kind, or customary tooth extraction, has been observed in any skull classified as that of a Tasmanian full-blood. Neither dental caries nor any other pathological condition was noted in skulls of Tasmanian full-bloods who lived in the natural state prior to contact with civilized people. The palate of the Tasmanian full-blood is wide and only moderately high, and in many cases it has a well-defined maxillary palatine torus. The teeth of the Tasmanians are smaller than those of the Australians.

My thanks are due to all who have permitted me to examine the Tasmanian remains which are in their charge, and to the University of Melbourne for financial assistance. I am particularly indebted to Dr G. M. Morant for reviewing and correcting the typescript and arranging it for publication; to Mr D. J. Mahony, M.Sc., Director of the National Museum, Melbourne, for revising the manuscript; to Dr E. Ford, Lecturer in Anatomy in the University of Melbourne, for checking the measurements of the limb bones; to Acting Professor M. H. Belz, of the University of Melbourne, for carrying out some of the calculations, and to Mr W. H. Preston for photographing the skulls.

REFERENCES

BERRY, R. J. A. & ROBERTSON, A W. D. (1909*a*). "Preliminary Communication on Fifty-three Tasmanian Crania, Forty-two of which are now recorded for the First Time." *Proc. Roy. Soc. Vict.* **22**, 47–58.

—— —— (1909*b*). "Preliminary Account of the Discovery of Forty-two hitherto unrecorded Tasmanian Crania." *Anat. Anz.* **35**, 11–17.

—— —— (1909*c*). "Dioptographic Tracings in Four Normae of Fifty-two Tasmanian Crania." *Trans. Roy. Soc. Vict.* **5**.

—— —— (1910). "The Place in Nature of the Tasmanian Aboriginal as deduced from a Study of his Calvaria. Part I." *Proc. Roy. Soc. Edin.* **31**, 41–69.

—— —— (1914) "The Place in Nature of the Tasmanian Aboriginal as deduced from a Study of his Calvaria. Part II" *Proc. Roy. Soc. Edin.* **34**, 144–89.

BERRY, R. J. A., ROBERTSON, A. W. D. & BÜCHNER, L. W. G. (1914). "The Craniometry of the Tasmanian Aboriginal." *Jour. Roy. Anthrop. Inst.* **44**, 122–6.

BERRY, R. J. A., ROBERTSON, A. W. D. & CROSS, K. S. (1910). "A Biometrical Study of the Relative Degree of Purity of Race of the Tasmanian, Australian, and Papuan." *Proc. Roy. Soc. Edin.* **31**, 17–40.

BÜCHNER, L. W. G. (1912*a*). "An Investigation of Fifty-two Tasmanian Crania by Klaatsch's Cranio-trigonometrical Methods." *Proc. Roy. Soc. Vict.* **25**.

—— (1912*b*). "A Study of the Prognathism of the Tasmanian Aboriginal." *Proc. Roy. Soc. Vict.* **25**.

CROSS, K. S. (1910). "On a Numerical Determination of the Relative Positions of certain Biological Types in the Evolutionary Scale, and of the Relative Values of Various Cranial Measurements as Criteria." *Proc. Roy. Soc. Edin.* **31**, 70–84.

DUCKWORTH, W. L. H. (1904). *Morphology and Anthropology.*

HARPER, W. R. & CLARKE, A. H. (1898). "Notes on the Measurements of the Tasmanian Crania in the Tasmanian Museum, Hobart." *Pap. and Proc. Roy. Soc. Tas. for* 1897. 97–110.

HRDLIČKA, A. (1928). "Catalogue of Human Crania in the United States National Museum Collections. Australians, Tasmanians, etc." *Proc. United States Nat. Mus.* **71**, 1–40.

LEWIS, A. N. (1934). "Correlation of the Tasmanian Pleistocene Raised Beaches and River Terraces in Unglaciated Areas." *Pap. and Proc. Roy. Soc. Tas. for* 1934.

LORD, C. E. (1919). "Preliminary Note upon the Discovery of a Number of Tasmanian Aboriginal Remains at Eaglehawk Neck." *Pap. and Proc. Roy. Soc. Tas. for* 1918.

LORD, C. E. & CROWTHER, W. L. (1920). "A Descriptive Catalogue of the Osteological Specimens relating to the Tasmanian Aborigines contained in the Tasmanian Museum." *Pap. and Proc. Roy. Soc. Tas. for* 1920.

MORANT, G. M. (1927). "A Study of the Australian and Tasmanian Skulls, based on previously published Measurements" *Biometrika*, **19**, 417–40.

SMITH, W. RAMSAY (1916). "A Description of some Tasmanian Skulls." *Rec. Aus. Mus.* (2), **11**.

TURNER, SIR WILLIAM (1884). "Report on the Human Crania and other Bones of the Skeletons collected during the Voyage of H.M.S. Challenger in the years 1873–1876." *Challenger Reports*, **10**.

—— (1908). "The Craniology, Racial Affinities and Descent of the Aborigines of Tasmania." *Trans. Roy. Soc. Edin.* (2), **46**, 365–403.

—— (1910). "The Aborigines of Tasmania. Part II. The Skeleton." *Trans. Roy. Soc. Edin.* (3), **47**, 411–54.

—— (1914). "The Aborigines of Tasmania. Part III. The Hair of the Head." *Trans. Roy. Soc. Edin.* (2), **50**, 309–47.

WEST, J. (1852). *The History of Tasmania*, **2**, 1–98.

WOOD JONES, F. (1929*a*). "Measurements and Landmarks in Physical Anthropology." *Bernice P. Bishop Mus., Bull.* **63**.

—— (1929*b*). "The Tasmanian Skull." *Jour. Anat.* **63**.

—— (1931). "The Non-metrical Morphological Characters of the Skull as Criteria for Racial Diagnosis." *Jour. Anat.* **65**.

WOOD JONES, F. & CAMPBELL, T. D. (1924). "Six hitherto undescribed Skulls of Tasmanian Natives." *Rec. South Aus. Mus.* **2**, 459–69.

WUNDERLY, J. (1935). "The Tasmanian Crania in Collections in the Commonwealth." *Med. Jour. Aus.* April 13, 1935, pp. 455–60.

WUNDERLY, J. & WOOD JONES, F. (1933). "The Non-metrical Morphological Characters of the Tasmanian Skull." *Jour. Anat.* **67**.

APPENDIX I

The Tasman series of skulls

At the commencement of the examination of specimens catalogued as Tasmanian in Commonwealth collections it was found to be impossible to identify readily a majority of them from the descriptions previously published, because of unsystematic and individual methods of numbering. It was decided, therefore, to label each skull "Tasman" series with a new serial number. The numbers from 1 to 52 in this series are the same as those allotted by Berry & Robertson (1909*a*). The Tasman series is made up by the remains of 114 adult and juvenile individuals represented by complete or incomplete skulls, and in some cases by mandibles only. The following particulars are given in the lists and folding table (Appendix III) below.

(i) The collection in which each specimen is preserved at present and its number in this collection. The abbreviations used are: *Public collections.* A.M.S. = Australian Museum, Sydney; I.A.C. = Institute of Anatomy, Canberra; M.C.D. = Municipal Council, Devonport; N.M.M. = National Museum, Melbourne; Q.V.M. = Queen Victoria Museum, Launceston; S.A.M. = South Australian Museum, Adelaide; T.M.H. = Tasmanian Museum, Hobart; U.M. = University of Melbourne. *Private collections.* A.L.M. = A. L. Meston, Esq., M.A., Launceston; G.R. = Gilbert Rigg, Esq., Melbourne; W.I.C. = Dr W. I. Clark, Hobart; W.L.C. = Dr W. L. Crowther, Hobart; H.A. = Howard Amos, Esq., Cranbrook, Tasmania.

(ii) The number allotted to each specimen in previously published papers. The sources are given in the references above and the abbreviations denoting authors used are: B. &

R. = Berry & Robertson; H. & C. = Harper & Clarke; H. = Hrdlička; S. = Ramsay Smith; W. J. & C. = Wood Jones & Campbell.

(iii) The sexes of the specimens. The sexes given are those decided on by the writer after examination of the skulls, and unless otherwise indicated these are the same as those adopted by the earlier investigators of the material. Juvenile skulls and isolated mandibles are not sexed. Remarks on sexing are given on p. 319 above.

The essential aim of the enquiry described in the present paper was to distinguish between the skulls which are, and those which have been alleged to be but which in fact are not, remains of full-blood Tasmanians. This question is discussed fully in the text above. The following eight groups were distinguished and all the 114 specimens are assigned to one or other of these:

Section A. Tasmanian full-blood: sixty seven specimens. Particulars and measurements of the thirty-one adult male and twenty-seven adult female skulls assigned to this group are given in the table of individual measurements (Appendix III). The following specimens are also included in it:

No. in Tasman series	Description	Museum No.	Collection
82	Juvenile skull	A. 559	T.M.H.
90	"	L. 1/119	"
70	Mandible	A. 16541	S.A.M.
84	"	A. 2878	T.M.H.
85	"	A. 2214	"
97	"	A. 580	"
98	"	A. 2210	"
100	"	23389B	N.M.M.
103	"	—	A.L.M.

In the present state of our knowledge, it is not possible to distinguish at all accurately between mandibles of Tasmanian full-bloods and those of Tasmanian-Australian mixed-bloods, and hence the allocation of these eight mandibles to the Tasmanian full-blood group is particularly uncertain.

Section B. Australian full-blood: twelve specimens.

No. in Tasman series	Description	Museum No.	Collection	No. in earlier papers
25	Adult ♂ skull	1201	Q.V.M.	B. & R. 25
59	"	T (*d*)	I.A.C.	—
112	"	19	W.L.C.	—
113	"	—	Q.V.M.	—
49	Adult ♀ skull	12	W.L.C.	B. & R. 49
51	"	12922	N.M.M.	B. & R. 51
63	"	A. 577	S.A.M.	W. J. & C. 577; H., A. 577
77	"	A. 1649	T.M.H.	—
87	"	15	W.L.C.	—
88	"	16	"	—
68	Juvenile skull	A. 16539	S.A.M.	—
64	Mandible	A. 707	"	—

The three of these skulls described by Berry & Robertson were accepted as Tasmanian by them. No. 63 was accepted as Tasmanian by Wood Jones & Campbell, and Hrdlička

describes it as being of Australian type. Nos. 63 and 68 were found on the west coast of Tasmania.

Section C. A male skull (Tasman series No. 114) in the National Museum, Melbourne, which is apparently that of an individual who had no Tasmanian or Australian ancestors.

Section D. Tasmanian-European mixed-blood: seven skulls.

No. in Tasman series	Sex	Museum No.	Collection	No. in earlier papers
56	♂	A. 2228	T.M.H.	—
11	♀	4292	,,	H. & C. 12; B. & R. 11
14	♀	4290	,,	H. & C. 3A; B. & R. 14
15	♀	4295	,,	B. & R. 15
16	♀	4296	,,	B. & R. 16, ♂
52	♀	12997A	N.M.M.	B. & R. 52
105	♀	—	H.A.	—

Five of these skulls described by Berry & Robertson were accepted as Tasmanian by them. No. 11 was accepted as Tasmanian by Harper & Clarke and they group No. 12 as half-caste.

Section E. Australian-European mixed-blood: three female skulls.

No. in Tasman series	Museum No.	Collection	No. in earlier papers
55	23389	N.M.M.	—
96	1221	Q.V.M.	—
101	B. 3496	A.M.S.	S., A, ♂

No. 101 was accepted as Tasmanian by Ramsay Smith.

Section F. Tasmanian-Australian mixed-blood: nine skulls. Particulars and measurements of eight of these are given in Appendix III. The other is a male specimen (Tasman series No. 71) in the Tasmanian Museum, Hobart, where it is numbered 11509. Five of the skulls (Nos. 54, 65, 66, 67 and 109) were found at the northern end of the west coast, and three others (Nos. 29, 30 and 31) about 80 miles distant on the north coast.

Section G. Apparently skulls of individuals of mixed blood with no Tasmanian or Australian ancestry.

No. in Tasman series	Sex	Museum No.	Collection	No. in earlier papers
13	♂	4297	T.M.H.	H. & C. 2A; B. & R. 13
12	♀	4302	,,	H. & C. 1A, ♂; B. & R. 12 ♂

According to Harper & Clarke both these skulls are those of half-castes, and Berry & Robertson accepted them as Tasmanian.

Section H. Skulls which have been lost, or which are too fragmentary to yield reliable data: thirteen specimens.

No. in Tasman series	Museum No.	Collection	No. in earlier papers
21	1572	T.M.H.	B. & R. 21, ♀
22	A. 506	"	B. & R. 22, ♀?
23	A. 507	"	B. & R. 23, ♂?
24	—	(Lost)	B. & R. 24, ♂
39	—	(Lost)	B. & R. 39, ♂
41	4	W.L.C.	B. & R. 41, ♀
47	10	"	B. & R. 47, ♀
48	11	"	B. & R. 48
60	T (*n*)	I.A.C.	—
72	11554	T.M.H.	—
99	D. 607	"	—
104	1254	A.M.S.	S., C
108	T.M. 1644	T.M.H.	—

Seven of the eight of these skulls described by Berry & Robertson are accepted by them as Tasmanian, the other (No. 48) being supposed half-caste if not Tasmanian. No. 104 is classed as "Tasmanian?" by Ramsay Smith. No. 3 in Harper & Clarke's list is lost and it is not included in the Tasman series.

The cranial measurements in Appendix III were obtained by following the definitions of the International (Monaco) Agreement of 1906, a translation of the report being given by Hrdlička in his *Anthropometry* (1920). The numbers of this list are given and also the letters denoting the measurements customarily employed in craniological papers in *Biometrika*. The additional measurements of the "minimum" and "maximum" thickness of the left parietal were obtained by following Hrdlička's instructions (*op. cit.* p. 107). These are:

"Introduce one branch of compass into the cranial cavity, apply to anterior part of the lower portion of the parietal approximately 1 cm. above the squamous suture, bring other branch in contact with the bone externally, and pass backwards at about the same distance from the sutures, watching the scale of the instrument. Record observed minimum and maximum."

The cranial capacity was determined with fine spherical seed and this was packed as tightly in the skull as in the glass measuring cylinder, as far as could be told.

APPENDIX II

Limb bones found at Eaglehawk Neck

An account of the discovery of this material and a description of the skulls are given in Section 5 of the text. Considering the scarcity of Tasmanian limb bones, the discovery of a number of them at Eaglehawk Neck is of importance. Nine femora and six tibiae are sufficiently well preserved to provide reliable data. Unfortunately it is impossible to identify, with certainty, any two or more bones as having belonged to one individual. For this reason, the femora were given numbers and the tibiae letters. The bones of the upper limb were badly damaged at the time of unearthing, and are unsuitable for reliable measurements.

The measurements of the femora and tibiae, hitherto unrecorded, were made in accordance with the directions supplied by Wood Jones (1929*a*). They are:

Femur. 1. Maximum length. 2. Oblique length. 3. Maximum trochanteric length.

4. Oblique trochanteric length. 5. Antero-posterior diameter. 6. Lateral diameter. 7. Circumference of shaft. 8. Subtrochanteric transverse diameter. 9. Subtrochanteric antero-posterior diameter. 10. Maximum diameter of articular surface of head. 11. Minimum diameter of articular surface of head. 12. Epicondylar breadth. 13. Condylar breadth.

Tibia. 1. Maximum length. 2. Direct length. 3. Axial length. 4. Breadth of the condyles. 5. Antero-posterior diameter of shaft. 6. Transverse diameter of shaft. 7. Transverse cnemic diameter. 8. Sagittal cnemic diameter. 9. Antero-posterior diameter at level of tuberosity. 10. Transverse diameter at level of tuberosity. 11. Minimum circumference of shaft.

Measurements of femora

Serial No.	Museum No.	No. of measurement												
		1	2	3	4	5	6	7	8	9	10	11	12	13
1	A(E.H.) 756	480	478	467	455	33	31	102	42	31	48	42	84	78
2	762	470	466	458	445	31	27	93	37	26	45	42	82	78
3	757	459	456	445	437	32	28	94	38	30	46	41	78	73
4	763	457	455	444	435	33	30	96	38	29	47	44	79	—
5	759	436	435	430	425	31	27	90	38	31	43	40	74	—
6	758	415	413	406	395	28	22	81	35	25	41	—	72	68
7	761	491	487	475	—	35	31	105	43	31	48	44	84	80
8	766	—	—	—	—	28	23	80	37	26	40	—	—	—
9	764	—	—	—	—	31	28	90	40	32	43	40	—	—
	No.	7	7	7	6	9	9	9	9	9	9	7	7	5
	Mean	458·3	455·7	446·4	432·0	31·3	27·4	92·3	38·7	29·0	44·6	41·9	79 0	75·4

Serial No.	Length index (7/1)	Thickness index (5/6)	Platymeric index (9/8)	Head-length index (10/1)	Condyle-length index (13/1)
1	21·3	106·5	73·8	10·0	16·3
2	19·8	114·8	70·3	9·6	16·6
3	20·5	114·3	78·9	10·0	15·9
4	21·0	110·0	76·3	10·3	—
5	20·6	114·8	81·6	9·9	—
6	19·5	127·3	71·4	9·9	16·4
7	21·4	112·9	72·1	9·8	16·3
8	—	121·7	70·3	—	—
9	—	110·7	80·0	—	—
No.	7	9	9	7	5
Mean	20·6	114·8	75·0	9·9	16·3

manian full-bloods

$O_1'R$. 16	$O_1'L$: 16	O_2R: 17	O_2L· 17	NH: 13	NB: 14	Alv. B:18	Alv. arch. H.18a	G_2: ? 19b	G_1': 19a	100 B'/B	100 O_2/O_1', L	100 G_2/G_1'	100 fmb'/fml
38	37	33	33	48	27	68	62	38	51	69·1	89 2	74·5	[illegible]
39	38	25	25	41	27	—	—	—	—	69·5	65·8	—	
37	36	27	27	42	26	—	—	—	46	65·8	75·0	—	81·3
38	38	28	28	44	28	—	—	—	—	—	73·7	—	
38	39	30	29	49	27	66	60	38	—	66·0	74·4	—	[illegible]
39	39	29	29	42	28	—	—	44	—	70·4	74·4	—	[illegible]
39	39	30	29	48	28	73	—	44	—	68·3	74·4	—	[illegible]
39	—	32·5	—	43	—	—	—	—	—	75·0	—	—	[illegible]
38	—	29	—	46	—	—	—	—	—	70·4	—	—	
38	38	30	30	46	27	64·5	—	40·5	—	66·2	78·9	—	[illegible]
38	38	34	33	50	30	66	61	40	—	70·8	86·8	—	[illegible]
38	39	30	29	44	30	—	—	—	—	71·2	74·4	—	[illegible]
—	38	—	30	—	—	65	—	41	48	—	78·9	85·4	88·2
37	38	31	31	43	28	66	—	38	—	70·7	81·6	—	[illegible]
—	—	—	—	—	—	—	—	—	—	66·9	—	—	
36 5	36·5	30·5	30·5	48·5	26·5	67 5	—	38·5	52	69·9	83·6	74·0	94·2
—	—	—	—	—	—	—	—	—	—	—	—	—	[illegible]
—	36	—	28	41	25·5	62	63	34	53	—	77·8	64·2	
42	—	33	—	—	—	—	—	—	—	67·9	—	—	70·0
—	—	—	—	—	—	—	—	—	—	71·9	—	—	
—	—	—	—	—	—	—	—	—	—	—	—	—	
—	—	—	—	—	—	—	—	—	—	—	—	—	
37	37·5	30	29·5	44	24	66·5	62	43	51	63·7	78·7	84·3	[illegible]
—	39	—	27	47	26	70	62	40	52	64·7	69·2	76·9	[illegible]
—	—	—	—	—	—	—	—	—	—	72·4	—	—	[illegible]
37	37	29	30·5	42	25	62	54	37	44	67·8	82·4	84·1	[illegible]
37	37	28·5	29	46	25	62	62	37	50	70·6	78·4	74·0	[illegible]
39	37	29	30	45	27	65	60	39	49	69·5	81·1	79·6	[illegible]
—	—	—	—	—	—	—	—	—	—	66·4	—	—	[illegible]
—	—	—	—	—	—	—	—	—	—	69·6	—	—	
—	—	—	—	47	27	—	—	—	—	—	—	—	[illegible]
40	38	31	29	50	29	71	64	45	52	70·6	76·3	86·5	[illegible]
—	38	—	33	49	27	—	—	—	—	65·2	86·8	—	101·8
—	—	—	—	—	—	—	—	—	—	67·2	—	—	
36	36	30	30	44	26	58	56	31	49	67·2	83·3	63·3	[illegible]
36	36	31	30	44	25	—	—	—	—	64·2	83·3	—	[illegible]
38	39	33	31	—	—	—	—	—	—	73·5	79·5	—	[illegible]
—	—	—	—	—	—	—	—	—	—	68·2	—	—	[illegible]
—	—	—	—	—	—	—	—	—	—	64·7	—	—	[illegible]
37	—	32	—	40	25	64	—	38	—	69·3	—	—	
38	—	32	31	50	27	66	—	—	—	—	—	—	[illegible]
—	—	—	—	—	—	—	—	—	—	70·1	—	—	[illegible]
37	37	29	29	42	24	64	59	36	51	66·9	78·4	70·6	[illegible]
36	35	27	27	45	27	—	—	—	—	77·6	77·1	—	[illegible]
37	37	29	31	42	28	61	—	36	—	71·0	83·8	—	
37	37	32	31	48	28	—	—	—	—	70·8	83·8	—	[illegible]
37	—	35	33	42	27	—	—	—	—	71·0	—	—	[illegible]
35	35	29	29	44	25	66	54	39.	47	—	82·9	83·0	
—	—	—	—	—	—	—	—	—	—	71·0	—	—	[illegible]
37	37	31·5	31 5	44·5	25	63	60	36	52	65·8	85·1	69·2	[illegible]
38	39	35	34	50	26	—	58	—	48	65·0	87·2	—	[illegible]
33	32	28·5	29	37 5	24·5	59	51	33	—	67·4	90·6	—	
35	35 5	30	31	44·5	26	62·5	57	36·5	47·5	66·9	87·3	76·8	[illegible]
—	—	—	—	—	—	—	—	—	—	69·6	—	—	[illegible]
38·5	38	31	30	43	25·5	66	57	38 5	46·5	68·4	78·9	82·8	[illegible]
37	39	30	30	47	26	68	59	40	49	71·4	76·9	81·6	[illegible]
37	38	28·5	29	42	24	63	53	40	45	67·2	76·3	88·9	[illegible]
39	39	34	34	—	22·5	63	—	40	49	67·4	87·2	81·6	[illegible]

is found in the domain of the west coast tribe

38	36	31	31	49	27	71	63	42	56	72·5	86·1	75·0	[illegible]
37	37	30 5	30 5	51	30	67	—	39	49	69·9	82·4	79·6	[illegible]
40	—	34	—	—	—	—	—	—	—		—	—	
42	43	31	33	51	27·5	72	66	41	59	66·9	76·7	69·5	[illegible]
40	42	31	30	57	22·5	69	61	41	54	66·4	71·4	75·9	[illegible]
38	38	32	32	52	28	70	61	44	52	81·6	84·2	84·6	[illegible]
38	37	31	31	48	27	65	58	39	49	71·4	83·8	79·6	
39	40	34	34	51	29	68	63	41	54	63·9	85·0	75·9	[illegible]

1 as Tasmanian.

‡ Wood Jones & Campbell accept Nos. 6

Measurements of tibiae

Serial letter	Museum No.	No. of measurement										
		1	2	3	4	5	6	7	8	9	10	11
A	A (E.H.) 794	366	357	351	68	31	24	25	36	41	33	81
B	797	393	389	377	—	35	22	25	42	45	34	85
C	793	—	—	—	—	40	24	—	—	55	—	90
D	795	388	380	369	78	33	20	23	39	46	32	81
E	798	355	352	342	70	31	19	24	37	43	32	71
F	796	—	—	—	—	35	23	—	—	—	—	85
No.		4	4	4	3	6	6	4	4	5	4	6
Mean		375·5	369·5	359·7	72·0	34·2	22·0	24·2	38·5	46·0	32·7	82·2

Serial letter	Length index (11/1)	Thickness index (6/5)	Cnemic index (7/8)
A	22·1	77·4	69·4
B	21·6	62 9	59·5
C	—	60·0	—
D	20·9	60·6	59·0
E	20·0	61·3	64·9
F	—	65·7	—
No.	4	6	4
Mean	21·2	64·6	63·2

Five of the nine unsexed femora are hyperplatymeric (74·9 and under) and the other four are platymeric. Two of the tibiae are platycnemic (55·0–62·9) and the other two mesocnemic. Turner (1910) states that the stature of the Tasmanians, "as determined by measurements made during life", ranged "in men from 5 ft. 1 in. to 5 ft. 6 or 7 in., with a mean of 5 ft. 3¾ in., and in women from 4 ft. 3 in. to 5 ft. 4 in., with the mean 4 ft. 11¼ in." Using the formula to which he refers—stature = 2 (oblique length of femur + condylo-astragalar length of tibia) + 26 mm.—the mean measurements of the unsexed Eaglehawk Neck femora and tibiae give a mean stature of 5 ft. 5 in.

NOTE ON Dr J. WUNDERLY'S SURVEY OF TASMANIAN CRANIA

By G. M. MORANT

There can be little doubt that some of the alleged Tasmanian skulls in Commonwealth collections for which data have been published are not of pure Tasmanian origin, and anthropologists are indebted to Dr Wunderly for having made a comprehensive and careful enquiry into the authenticity of each specimen. He explains that a number should be rejected, partly because there are no adequate records to authenticate them, but principally on account of the fact that their characters distinguish them from the genuine crania. One may accept his diagnosis as correct in the majority of cases, at least, and yet remember the danger that anatomical selection of a racial group may lead to a sample with unnaturally small variability. An examination of any random series of skulls which correctly represents a specialized racial population—such as the Guanche, the Andamanese or the Greenland Eskimo—shows that a number of the individuals included may depart quite markedly from the type for the series.

Dr Wunderly's measurements of "full-blood Tasmanian" crania may be compared with those given previously for a sample which almost certainly includes some spurious specimens. The series made up by the present writer (1927) by pooling data given by a number of anthropologists may be used for this purpose. It includes more than half of Dr Wunderly's accepted sample, a few specimens rejected by him, and a number in European collections—some of which may be unauthentic—which he has not measured. In comparing constants for these two groups it must be remembered that personal equation in measuring and changes in the estimates of sex may be partly responsible for the differences observed, while sampling errors are large as both series are small. Means are given in Table I on p. 321 above. The differences between corresponding pairs are nearly all of the order expected for samples of the sizes available drawn from the same population, and the agreement in the case of the indices is particularly close. A bad agreement is only found in the case of the interorbital breadth, and it is extremely probable that this is due to the fact that different definitions of the measurement were used. If it is ignored, male and female comparisons can be made for twenty-two absolute measurements. For these Wunderly's female mean exceeds Morant's in eighteen cases, but the same tendency is not observed in the case of the male series. It can be seen from Dr Wunderly's table of individual measurements that in his "full-blood Tasmanian" series the fifteen male skulls previously described had all been classed as male by the earlier anthropologists, but of his nineteen female nine had previously been classed as male. The transference of an appreciable proportion of specimens from the male to the female group will be expected to lower the male and raise the female means. Only one of these effects is observed, but in view of the close agreement of the indices it appears probable that the resexing of the material has had more effect on the means of absolute measurements than the rejection of doubtful specimens.

A few standard deviations (with their probable errors) are given in the table below. It is customary to find that these constants tend to be slightly less for a female than for a male series representing the same population in the case of absolute measurements, and to be approximately the same for the two sexes in the case of indices. Wunderly's male and female series are quite unexceptional in this respect, but when compared with Morant's their standard deviations are seen to be appreciably less in the case of the maximum calvarial breadth and cephalic index.

	Male		Female
	Tasman series (Wunderly)*	Pooled "Tasmanian" series (Morant)	Tasman series (Wunderly)*
Maximum length (L)	7·58 ± 0·66 (30)	6·01 ± 0·44 (43)	5·94 ± 0·57 (25)
Glabella-inion length	8·12 ± 0·71 (30)	—	5·77 ± 0·55 (25)
Maximum breadth (B)	4·11 ± 0 38 (27)	5·32 ± 0·33 (60)	4·62 ± 0·44 (25)
Maximum frontal breadth (B'')	4·36 ± 0·42 (25)	—	4·36 ± 0·42 (25)
Bimastoid breadth	5·42 ± 0·51 (26)	—	—
Minimum frontal breadth (B')	4·35 ± 0·41 (25)	4·81 ± 0·29 (62)	4·15 ± 0·38 (27)
Basio-bregmatic height (H')	5·11 ± 0·50 (24)	4·76 ± 0·31 (55)	—
Auriculo-bregmatic height	3 72 ± 0·35 (25)	—	—
Horizontal circumference (U)	16·16 ± 1 5 (25)	14·10 ± 0·97 (48)	—
Arc nasion to bregma (S_1)	6·75 ± 0·62 (27)	5·98 ± 0·43 (44)	5·07 ± 0·47 (26)
Arc bregma to lambda (S_2)	5 71 ± 0·51 (28)	6·77 ± 0·50 (42)	5·59 ± 0·54 (24)
Broca's transverse arc	11·06 ± 1·1 (25)	10·92 ± 0·82 (40)	—
Interorbital breadth	2·56 ± 0·24 (25)	—	—
100 B/L	2·14 ± 0·20 (27)	2·58 ± 0·19 (43)	1·66 ± 0·16 (24)
100 H'/L	2·39 ± 0·23 (24)	2·21 ± 0·17 (37)	—
100 B'/B	2·71 ± 0·26 (24)	—	3·00 ± 0·29 (25)

* Constants provided by Dr Wunderly.

The following distributions are for the latter character. The samples are too small to yield any decisive conclusions, but there is certainly a suggestion that the female distribution for Dr Wunderly's measurements has been curtailed. The standard deviation for it is appreciably lower than that recorded for an unselected series of skulls from any part of the world.

Cephalic index (central values)	68·5	69·5	70·5	71·5	72·5	73·5	74·5	75·5	76·5	77·5	78·5	79·5	80·5	Total
♂ Wunderly	—	—	1	2	5	8	3	2	1	4	0	1	—	27
♂ Morant	1	2	3·5	0·5	4	8	9·5	7·5	0	1·5	4	1·5	—	43
♀ Wunderly	—	—	—	—	—	3	1·5	4	6	6·5	2	1	—	24
♀ Morant	1	0	0	3	0·5	1·5	3	2	2	3	2	0	1	19

In my paper giving the means of the pooled series of Tasmanian skulls the coefficient of racial likeness with an Australian (the A) series is provided and it may be asked how such a comparison is affected by the revision of the material. Of the thirty-one characters used when possible for this purpose, there are twenty-one available for all three male series and these give the following coefficients. The Australian A standard deviations were used in these comparisons, though these are probably rather greater than the true values for the Tasmanian population, and hence all the coefficients are probably somewhat lower than they should be.

	Crude C.R.L.	Reduced C.R.L.
Tasmanian: Wunderly ($\bar{n} = 21·7$) and Tasmanian: Morant (48·3)	0·66 ± 0·21	2·21 ± 0·69
Tasmanian: Wunderly and Australian A (113·2)	11·34 ± 0·21	31·14 ± 0·57
Tasmanian: Morant and Australian A	13·72 ± 0·21	20·27 ± 0·30

The first of these comparisons is not really justifiable, since a certain number of specimens is common to the two Tasmanian series, but it shows that they have very similar mean measurements. At the same time, judging by the reduced coefficient, the selected series is distinctly further removed than the other from the Australian *A* series. The same situation is observed for all the characters considered singly which show significant differences between the Australian mean, on the one hand, and both Tasmanian means, on the other, with the single exception of the cephalic index. These means are:

	100 *B/L*	100 *B/H'*	100 *NB/NH*	*B*
Tasmanian (Wunderly)	74·2 (27)	106·2 (22)	59·9 (19)	138·2 (27)
Tasmanian (Morant)	74·2 (43)	103·9 (35)	59·1 (57)	136·0 (60)
Australian *A*	70·8 (94)	99·3 (156)	54·6 (132)	132·2 (162)

	LB	*G'H*	O_2	*NH*
Tasmanian (Wunderly)	98·1 (22)	62·4 (12)	29·9 (19)	45·1 (21)
Tasmanian (Morant	98·8 (55)	62·5 (36)	31·1 (60)	47·1 (58)
Australian *A*	102·1 (137)	66·8 (79)	33·5 (118)	49·5 (118)

Dr Wunderly's selection of the skulls has thus had the effect of modifying our conception of the Tasmanian type and making it still less like the Australian. The revision doubtless gives a closer approximation to the truth.

A

B

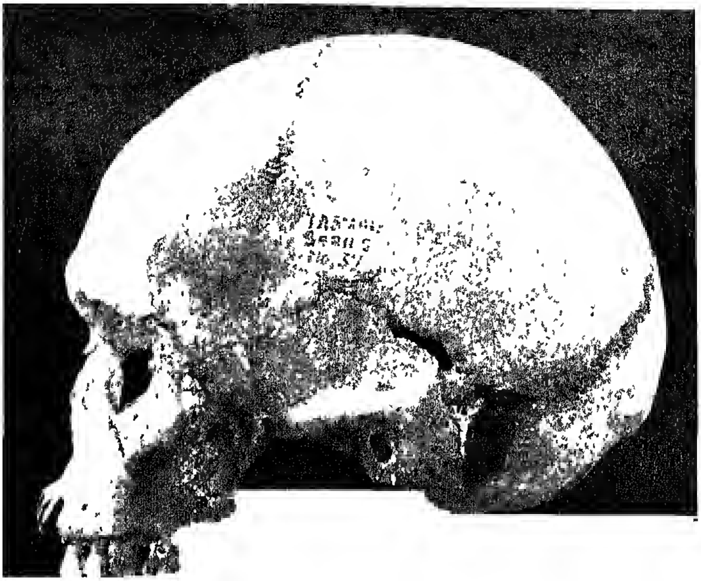
C

A typical female Tasmanian skull: Tasman series No. 38.

A

B

Typical male Tasmanian skulls: A, Tasman series No. 35; B, No. 36.

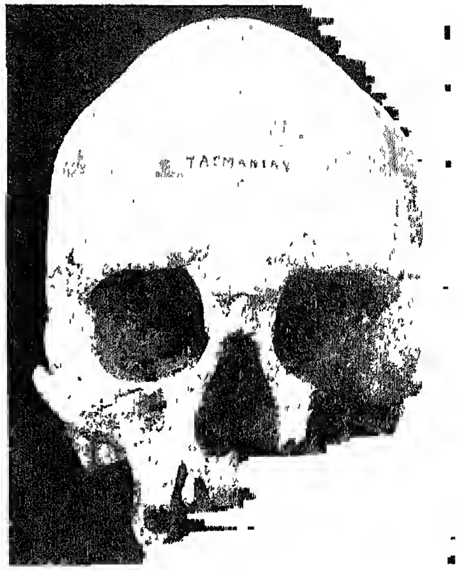


A

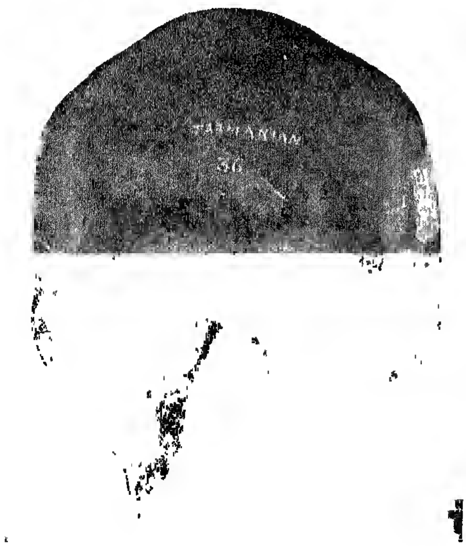
B

Typical female Tasmanian skulls: A, Tasman series No. 37; B, No. 34.

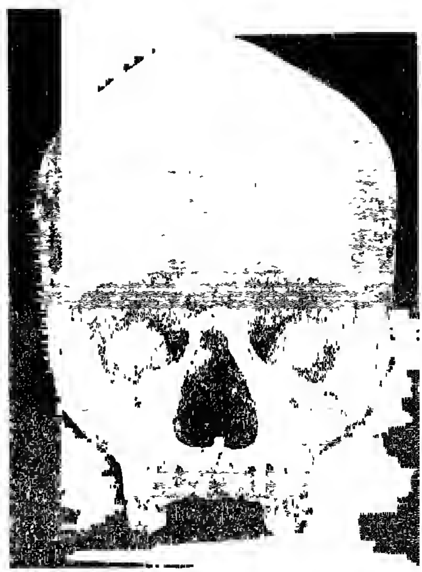
A



B



C

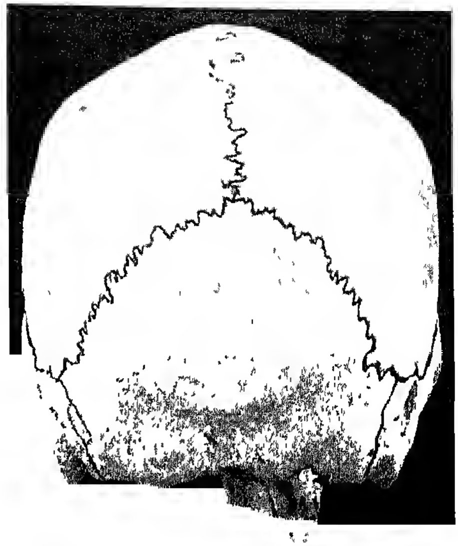


D

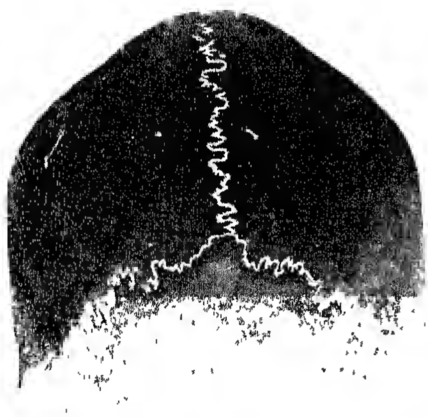
Typical Tasmanian skulls A, Tasman series No 34, female; B, No. 35, male, C, No. 36, male; D, No 37, female.

Wunderly: *Cranial and other Skeletal Remains of Tasmanians*

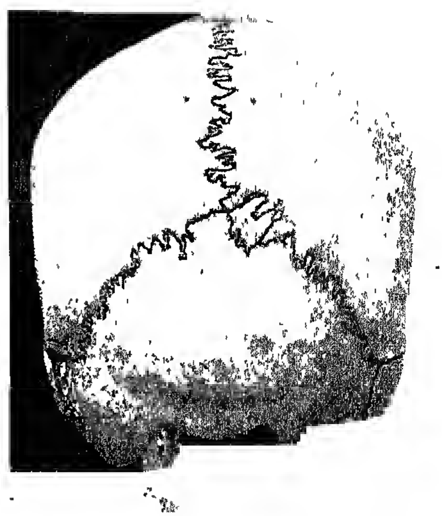
A



B



C



D

Typical Tasmanian skulls. A, Tasman series No 34, female; B, No 35, male; C, No 36, male, D, No 37, female

Male skull presumed to be of Tasmanian-Australian mixed-blood origin: A and B, Tasman series No. 65; C, D and E, No. 109; F, No 30.

DISEASE AND ENVIRONMENT

By E. A. CHEESEMAN, W. J. MARTIN AND W. T. RUSSELL
Of the Medical Research Council's Statistical Staff

From the Division of Epidemiology and Vital Statistics,
London School of Hygiene and Tropical Medicine

A CHARACTERISTIC feature of vital statistics is the greater mortality in town than in country. Why should this be? The most obvious difference is the closer contact between human beings. It may indeed be true that in some villages domestic overcrowding is as great as in towns, but the number of persons per acre is greatly less and the occupations of town-dwellers, whether in offices, shops or factories, involve a longer continued and more intimate contact of human being with human being than is the rule in country life.

The conditions of town life (even if we do not reckon the intense, temporary overcrowding, consequent upon transport from suburbs to centre, a very important factor of modern urban life) are evidently favourable to the droplet infection which, in so many diseases, is believed to be a principal means of transmission. Hence the relation between density of population and the incidence of disease and death must always be worthy of close study.

It is not a region hitherto unexplored but, in most of the early investigations, the effect of density has been measured solely in terms of mortality. This index is not quite satisfactory because before 1911 no transference of death was made to place of residence, and hence there must have been many instances in which there was an appreciable difference between the ostensible and real mortality of districts containing hospitals and institutions. In the present paper the work of earlier investigators has been reviewed and an attempt made by the adoption of more recent statistics to assess the relationship, not only between density and mortality in general, but also between density and the morbidity from infectious disease.

PREVIOUS INVESTIGATIONS

Farr

The problem of density and health was first examined by Farr. In the appendix to his *Fifth Annual Report* (Farr, 1843) he endeavoured to show that there was a definite relationship between density and mortality which was described by an equation of the form

$$D = C\delta^n,$$

where D is the crude death rate and δ the density (number of persons per square

mile). He found in his examination of the statistics for the districts of London as then constituted that "the mortality did not increase as their density but as the 6th root of their density". Farr (1875, pp. xxiii–xxiv) returned to the subject with an analysis of the data for 1861–70 and, in the *Decennial Supplement* for that period, gave the following account:

A larger basis is now supplied by the facts of the ten years recorded in all the districts of England and Wales. They have been arranged in the Tables; and with this result, that in every group the mortality increases with the density, but happily not in direct proportion of the density. London has been excluded in the following calculations. Thus in 345 districts with a mortality of 19·2 the density was 186 persons to a square mile; in 9 districts with a density of 4499 what was the mortality? In the first place it was not expressed by the proportion of 186:4499::19·2:x but by this proportion:

$$(186)^{\cdot 12}:(4499)^{\cdot 12}::19\cdot 2:x=28\cdot 1.$$

The observed and calculated rates as deduced by Farr for varying densities were:

Group of districts	Density persons per square mile	Crude death rate	
		Observed	Calculated
I	166	16·75	18·90
II	186	19·16	19·16
III	379	21·88	20·87
IV	1718	24·90	25·02
V	4499	28·08	28·08
VI	12357	32·49	32·70
VII	65823	38·62	38·74

He inserted the following footnote to this table (Farr, 1875, p. clviii):

m being the mortality in any group and m' being the higher mortality at any other group, D and D' being the density of population in the two groups then

$$m' = m\left(\frac{D'}{D}\right)^n = m\left(\frac{D'}{D}\right)^{\cdot 11998}$$

The mortality of the districts is nearly as the 0·12th root of their densities or taking the above value of n, and p and p' as the mean proximity of person to person we have

$$m' = m\left(\frac{p}{p'}\right)^{2n}.$$

So the mortality of the district is nearly as the 6th root of the proximities.

This statement is not arithmetically correct, as according to the value of n obtained by Farr the mortality varies as the 0·12*th power* and not as the 0·12*th root* of the density.

Ogle and Tatham

Neither of Farr's immediate successors, Ogle and Tatham, shared his belief that mortality was purely a function of density of population. Tatham (1895,

p. xlvi), discussing the density and mortality figures for 1881–90 in relation to Farr's "law", wrote:

although density and mortality generally increase or decrease together the relation between them is now too complex to admit of being expressed by a formula similar to that alluded to above.

Brownlee

For Brownlee (1922), Farr's equation held a considerable fascination, and the following quotations are a striking testimony of his belief in the soundness of the conception and its application to vital statistics.

His (Farr's) treatment of it is one of the brilliant attempts to extract the real meaning of figures so frequent in his work, but though this theory has not shared in the complete neglect that has been the lot of his attempt to put a quantitative measure to the course of epidemics it has suffered as much from the kind of patronage with which it is usually discussed.

He revived interest in the law and demonstrated its applicability to domains other than public health. In its relation to density and disease he stressed the fact that owing to wide variability in the mortality of districts possessing the same density of population the law can really only be a law of average. In his view

the effect of density is not merely as density. The country preserves life even in the presence of excess or dissipation: the town does not. Further, in the period of growth, children in the city do not get anything like the same chance as their fellows in the country, even though housing may be better and food more abundant. In addition, filth in the country is, at its worst, in most cases but a local nuisance spreading enteric and diarrhoea at times, but not having the power of rendering a whole district foetid. All these influences act concurrently and cumulatively to depress health the more closely people are crowded together, and as life is a physico-chemical process this effect must be measurable and should be capable of expression in some formula which goes back to chemistry and physics. Such a formula is that of Dr Farr.

The necessity of applying an equation of this nature to describe the relationship between density and mortality statistics was demonstrated by Brownlee in the statistics for Glasgow for the years 1898–1902.

Group of districts	Population 1901	Room density	Scarlet fever		Enteric fever	
			Cases	C.F.	Cases	C.F.
I	34868	0·5 –1·0	864	2·3	106	9·4
II	83255	1·0 –1·5	2148	3·5	389	12·8
III	201098	1·5 –2·0	5439	4·1	1308	15·8
IV	87885	2·0 –2·25	2184	5·0	711	16·3
V	237161	2·25–2·5	5610	5·1	1743	16·9
VI	117445	2·5 –2·75	2091	5·6	1003	16·3

The significance of the lesson conveyed by the trend of these case fatality rates is fairly obvious. Taking first the statistics for enteric, it is seen that, had

an investigation on the influence of insanitary conditions been confined to those localities in which the density value was over 1·5 persons per room, the inevitable conclusion would be that the environment of the person had no influence on the severity of the disease, as the fatality rates are nearly constant. Similarly for scarlet fever, but in this instance the upper limit of fatality is not reached till the concentration of the population is that of two persons per room.

Having been fully convinced that the relationship of density was best described by Farr's law, Brownlee then proceeded to apply the formula to the mortality data in the Registration Districts of England and Wales, grouped according to their densities for each decennial period since 1861–70. It will be remembered that Farr did similar calculations for the first decennium, but for his index of mortality he used the crude death rate. Brownlee would not accept either the crude or standardized rate as a suitable measure of ill-health. In his opinion the standardized death rate represented an impossible mortality in a stationary population—a standardized death rate of 13·49 per 1000 in the healthy districts would, in a stationary population, yield a mean life of

$$\frac{1000}{13{\cdot}49} \text{ or } 74 \text{ years,}$$

a figure, he said, hardly conceivable if the observed properties of life represent anything which is fundamental. But it is difficult to understand the reason why either the crude or standardized death rate should be dependent on what they represent in a stationary population. The crude death rate is a reality—the population has actually died at that rate—whereas the life table death rate, his preference, is governed by hypothetical considerations.

It may seem extraneous to our investigation to dwell at length on the appropriate measure of mortality which should be inserted in the formula, but to appreciate Brownlee's work it is necessary to do so. He definitely maintained that the life table death rate was the only satisfactory criterion of ill-health:

> it has one property which places it as a measure above either the standardized death rate or the crude death rate in as much as it has been found for England and Wales to be very closely connected with the density of population.

It must be pointed out that the life table death rates which he calculated for the various groups of districts were deducible from the standardized rates by means of linear equations. The final equations of the densities and the life table death rates in the Registration Districts of England and Wales as obtained by him for the successive decennial periods were:

$$\begin{aligned} &1861\text{–}70, && D = 12{\cdot}42\delta^{0{\cdot}1001}, \\ &1881\text{–}90, && D = 11{\cdot}45\delta^{0{\cdot}0985}, \\ &1891\text{–}1900, && D = 10{\cdot}83\delta^{0{\cdot}1008}, \\ &1901\text{–}10, && D = 9{\cdot}90\delta^{0{\cdot}1023}. \end{aligned}$$

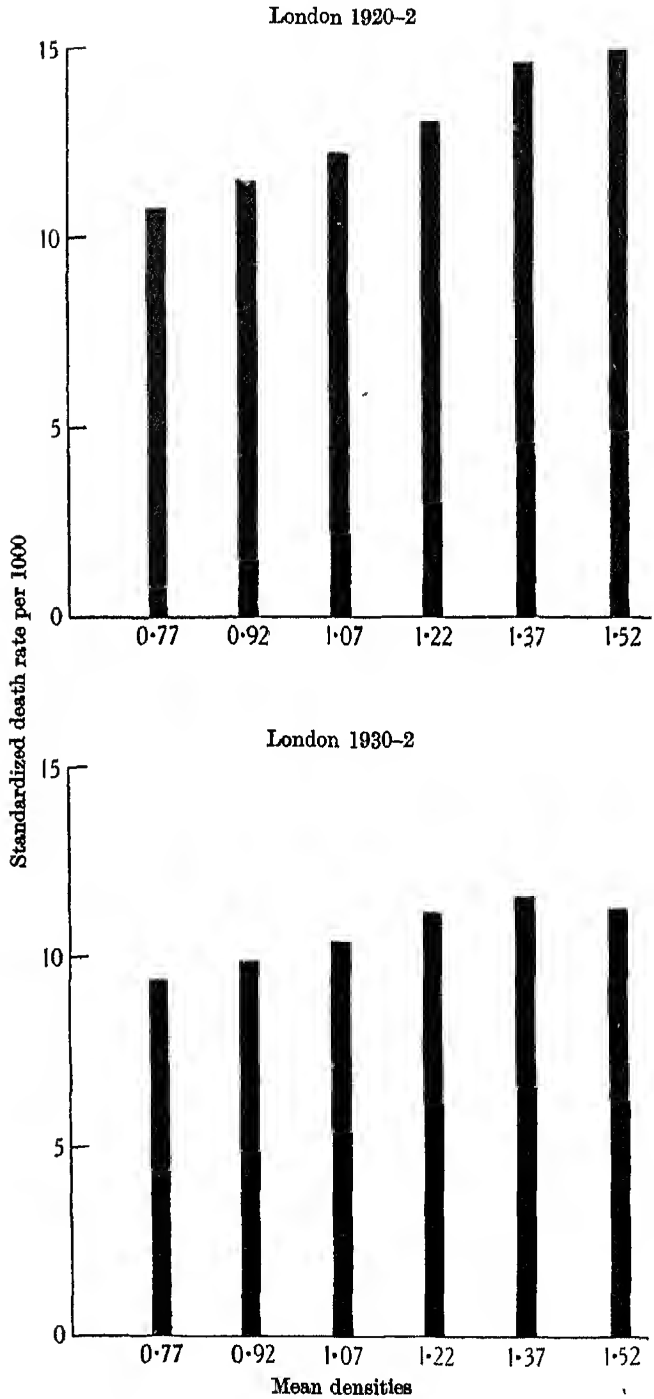
From these equations he concluded that:

> though general health has improved the power of the density has stood unchanged for forty years. That is to say that the death rate and the density remained related in essentially the samo way in the counties of England and Wales in 1905 as in 1865. It is the constant multiplier that has been affected by hygienic measures and not the law of the power. Hygiene acts surely all round but still is subjected to fundamental laws.

His adoption of the life table death rates in preference to either of the other measures invoked strong criticism.

In a discussion which took place at the Royal Statistical Society (1922) on "The Value of Life Tables in Statistical Research" the majority decried their use for this purpose and also opposed the theory that because it yielded a constant index in Farr's law the life table death rate possessed an intrinsic value. In the discussion the point was made that, since the life table death rate was deduced from the standardized death rate, by means of a linear relation, the relatively smaller variability of the life table death rate might be a mere artefact. This, however, does not impugn the conclusion that mortality, whatever the standard of measurement, does vary with density and that the relation is curvilinear. The death rates for 1920–2 and 1930–2 for London shown in the following diagram illustrate the point. The death rates for 1920–2 steadily increase over the range of densities, but the mortality during 1930–2 is at a maximum when the density is 1·37 persons per room and then declines slightly. This decline may be due to a random fluctuation arising from the smallness of the population living at the highest density group in 1930–2 as compared with 1920–2. Or possibly a saturation point is reached in 1930–2 at about 1·22 persons per room when, with increasing density, there is no *pro rata* increase in mortality. The last hypothesis could be reconciled with the results for 1920–2, since the saturation point will vary with the general level of the mortality rate. It would be higher in 1920–2 than in 1930–2 owing to the higher mortality rate in the former period, i.e. the saturation point in 1920–2 apparently occurs outside our range of densities. The results for 1930–2 give confirmation to an observation previously noted by Brownlee—that it is futile to conduct enquiries on the effects of environment on health and disease over a restricted range of environmental conditions. If such an enquiry had been conducted in London in 1930–2 and confined to districts having densities of 1·22 persons per room and over, then it is obvious that from a mortality viewpoint little or no differentiation would have arisen.

It seemed of interest to compare the relationship between density and mortality as described by a simple curve and a straight line respectively. For this purpose the standardized death rate was related to the two modern measurements of density, i.e. the number of persons per room and the percentage of the population living more than two to a room, and the crude death rate to the older


London 1920–2
London 1930–2
Standardized death rate per 1000
15
10
5
0
0·77
0·92
1·07
1·22
1·37
1·52
Mean densities

index, the number of persons per acre. The statistics which were utilized to effect the comparison were the death rates for the London boroughs for 1930–2, grouped into six classes according to density. The results are:

London, 1930–2

Mean no. of persons per room	Standardized death rates	$10{\cdot}19173\delta^{0\cdot3160}$	$2{\cdot}9181\delta+7{\cdot}2388$
0·77	9·34	9·38	9·49
0·92	9·87	9·93	9·92
1·07	10·34	10·41	10·36
1·22	11·14	10·85	10·80
1·37	11·56	11·26	11·24
1·52	11·23	11·63	11·67
Root mean square error		0·24	0·27
Mean % living more than two to a room	Standardized death rates	$0{\cdot}66869\delta^{0\cdot16960}$	$0{\cdot}11337\delta+8{\cdot}6626$
5·5	8·87	8·93	9·29
10·5	10·01	9·96	9·85
15·5	10·63	10·64	10·42
20·5	11·31	11·16	10·99
25·5	11·73	11 58	11·55
30·5	11·67	11·94	12·12
Root mean square error		0·14	0·31
Mean no. of persons per acre	Crude death rates	$7{\cdot}694356\delta^{0\cdot10480}$	$0{\cdot}01670\delta+10{\cdot}6490$
22·5	10·52	10·66	11·02
47·5	11·64	11·53	11·44
72·5	12·26	12·05	11·86
97·5	12·50	12·43	12·28
122·5	12·78	12·74	12·69
147·5	12·71	12·99	13·11
Root mean square error		0·16	0·33

Although the advantage lies in each case with Farr's expression, the advantage is not significant. Applying the z test to the corresponding estimates of

variance and combining the results, the improbability of such a concordance as observed is only moderate ($P=0\cdot078$). It must indeed be remembered that the method of fitting of Farr's equation (least squares applied to the logarithms) is not efficient, so that the superiority of the fit may be underestimated.

Any discussion on the effects of density on mortality would be very incomplete without mention of the painstaking studies by Stocks which are described in the Text or Part III of the *Annual Reports of the Registrar-General*, particularly for the year 1932. He investigated the influence not only of density but also of latitude on the general mortality and on specific diseases at all ages and at particular age periods. His main conclusion was:

> It seems fair to conclude that it is at these ages (1–5) that the greatest benefits may be anticipated as the overcrowding evil is mitigated.

It will be noted that in all these investigations, apart from that by Stocks, the effects of environment on mortality have been represented by the total death rate. *A priori* this criterion would seem to be too comprehensive. It is extremely unlikely that the child and the adult react in an equal degree to their environmental conditions, and hence the more appropriate examination would be one within specific age periods.

Childhood

There is no evidence that the unborn child is influenced by the mother's environment. The foetus has even been described as a true parasite protected against the vicissitudes of the mother and more or less independent even of her starvation or dissipation. But once the child begins its separate existence apart from its mother its immediate reaction to its surroundings will be best represented by a particular quota of the infant mortality. The reason for such discrimination is this. The death rate in the first year of life has been regarded as depending on three main factors: (1) shock of birth, (2) instability of the nervous and digestive system, (3) external factors embracing infection and environment.

If we accept the aggregated county boroughs and the aggregated rural districts as representing two widely divergent environments and then group their infantile deaths in age periods into two categories, A and B, in which A includes deaths of (1) and (2) character and B those in (3), thus:

A	B
Premature birth	Measles
Congenital malformation	Whooping cough
Congenital debility	Diarrhoea and enteritis
Injury at birth	Tuberculous diseases
Convulsions	Bronchitis and pneumonia

we obtain the following death rates and ratios for the decennium 1921–30:

	Age period in months			
	0–1	1–3	3–6	6–12
Group A:				
County Boroughs	26·94	4·98	2·07	1·42
Rural Districts	25·92	4·15	1·66	1·33
Ratio C.B./R.D.	1·04	1·20	1·25	1·07
Group B:				
County Boroughs	3·06	7·21	9·18	17·06
Rural Districts	1·96	4·52	4·83	8·70
Ratio C.B./R.D.	1·56	1·60	1·90	1·96

For the diseases in Group A, which may be regarded as non-preventable (deaths from prematurity accounting for the greater proportion), there is little difference between the rates in town and country in the first month of life, 4 %; between the ages of 1 month and 6 months there is greater divergence, 20–25 % but, for children aged 6–12 months, the mortality experience in the county boroughs is only 7 % worse than that in the rural districts. The comparison between the experience in town and country for diseases grouped under B is vastly different. The effects of an unhealthy environment are apparent in the first month of the child's post-natal existence and they become more accentuated with age. The mortality amongst children under 1 month in the county boroughs is 56 % higher than that in the rural districts, and for those aged 6–12 months the excess is no less than 96 %. In view of this disparity it is obvious that if further reduction in the national rate for infantile mortality is to be made the sanitarian or the public health administrator must concentrate his efforts to ameliorate the existing conditions in towns which help to produce the relatively high mortality from the diseases classified under Group B.

It must not be concluded, however, that the effect of environment is harmful only in infancy. The emphasis which has often been laid on the difference between the infant death rate in town and country would seemingly suggest this conclusion. It would not be valid. The influence of environment in childhood becomes progressively unfavourable and at age 2–3 it is at a maximum. This fact was clearly demonstrated by Brownlee. He selected certain life tables which related to the decennium 1891–1900 and which represented different environmental grades. He next expressed the mortality at ages in each as proportions of the rates at the corresponding ages in the Healthy District Life Table for the same period 1891–1900. We have chosen three life tables from his list:

Healthy district	=	Good environment
E 6 (England and Wales)	=	Average
Salford	=	Poor

The values are given in Table I.

TABLE I

Showing for the period 1891–1900 the ratio of the death rates in England and Wales (E 6), and in Salford, to those of the healthy districts (H 3) at individual ages between 0 and 5 years

Age	Death rates in healthy districts (H3)		Ratio			
			England and Wales (E6)		Salford	
	Males	Females	Males	Females	Males	Females
0	132·074	101·327	1·46	1·52	2·29	2·42
1	28·500	26·421	1·92	1·92	3·39	3·59
2	10·100	10·355	2·08	1·96	4·00	3·67
3	7·386	7·093	1·80	1·89	3·11	3·14
4	5·787	5·657	1·68	1·70	2·63	3·21
5	4·551	4·489	1·57	1·58	1·94	1·76

It will be seen that the maximum ratio in each instance is definitely at age 2–3 years.

To ascertain whether this phenomenon was a mere chance event or persisted before and after 1891–1900 we classified the registration counties into two groups, A, mainly urban, B, mainly rural, and calculated the appropriate ratios in the prescribed age periods. Subsequent to 1910 the ratios were based on

$$\left(\frac{\text{mortality in county boroughs}}{\text{mortality in rural districts}}\right).$$

The results are given in Table II and are of interest.

TABLE II

Showing the ratio of the mortality at individual ages under 4 years (1) in urban counties to that in rural counties, and (2) in county boroughs to that in rural districts

Areas	Period	Ratio							
		Males				Females			
		0–1	1–2	2–3	3–4	0–1	1–2	2–3	3–4
U.C./R.C.	1881–90	127	174	190	185	132	178	188	187
	1891–1900	131	187	203	190	137	194	202	194
	1901–10	132	199	204	184	138	203	209	188
C.B./R.D.	1911–14	138	388	374	346	142	397	406	349
	1920–23	133	211	184	—	135	206	189	—
	1930–32	135	189	165	151	134	182	171	142

Between 1881 and 1910 the ratio for males was highest at age 2–3 years. In 1911–14 a change occurred and the ratio at age 1–2 became the most important and has remained so up to the close of our experience. For females the regression did not take place until after the war. We can offer a more stringent test to show that the relative environmental influence of town and country on mortality is, at present, best indicated at age 1–2 years. In the Registrar-General's *Decennial Supplement*, 1931, Part I, life tables were made for the aggregated county boroughs in Northumberland and Durham, which we will call C, and for a group of rural districts in the Eastern counties (D). The q_x values of C expressed as proportions of D are:

	Age in years				
	0	1	2	3	4
Males C/D	1·66	3·77	3·03	2·67	2·19
Females C/D	1·64	3·54	2·74	2·48	1·99

These two areas C and D represent widely divergent types of environment and the maximum ratio is definitely at age 1–2 years.

It is now important to discover the particular diseases of childhood initially responsible for the highest ratio being in certain years. We are able to do this because in the *Decennial Supplement* for 1901–10, pp. ccxxiv–ccxxvi, the Registrar-General published the death rates from All Causes, specific diseases and groups of diseases in infancy and in the individual years of childhood up to age 5 during the period 1906–10 for two groups of counties, the one urban in character, the other rural. The rates and the ratio of the urban to the rural mortality are given in Table III.

It will be seen for All Causes of death that, although the ratio at age 2–3 years is still the largest, 1·97, it is really not much in excess of the value at the younger age. This is not unexpected, because we previously indicated that after 1911–14 the maximum index was definitely at age 1–2 years and the period 1906–10 is seemingly the transitional one: at least it borders that in which the regression occurred. The specific groups of diseases for which the ratio was definitely highest at age 2–3 years were the common infections, tuberculous, developmental and wasting.

Owing to the fact that deaths according to extent of urbanization and specific cause are no longer published for individual years of life between age 2 and 5 years (the existing age classification being 0–1, 1–2, 2–5) we are unable to indicate the diseases which produced the change in the age occurrence of the maximum ratio.

The statistics of the mortality in childhood which we have so far examined

TABLE III

Showing for the period 1906–10 *the deaths per* 1000 *survivors* (*both sexes*) *at the commencement of each year in England and Wales*

		Ages			
		1–2	2–3	3–4	4–5
I. Common infectious diseases	Urban	11·78	5·87	4·11	2·98
	Rural	4·88	2·36	2·01	1·73
	Ratio U./R.	241	249	204	172
II. Diarrhoeal diseases	Urban	5·03	0·83	0·29	0·16
	Rural	1·59	0·31	0·16	0·10
	Ratio U./R.	316	268	181	160
III. Developmental and wasting diseases	Urban	1·05	0·22	0·08	0·04
	Rural	0·65	0·13	0·07	0 03
	Ratio U./R.	161	169	114	133
IV. Tuberculous diseases	Urban	3·93	2·10	1·34	1·03
	Rural	2·14	1·03	0·71	0·66
	Ratio U./R.	184	204	189	156
V. Miscellaneous diseases	Urban	19·42	7·58	4·40	3·21
	Rural	11·74	4·61	2·90	2·24
	Ratio U./R.	165	164	152	195
VI. All causes	Urban	41·21	16·60	10·22	7·42
	Rural	21·00	8·44	5·85	4·76
	Ratio U./R.	196	197	175	156

for town and country support the viewpoint that the differences can be explained on the basis of earlier infection in the former. The pre-school child, as a consequence of his environment, is infected at an age when he is least able to resist a fatal attack. In rural districts infection occurs at a later age. Picken (1921) demonstrated this fact in connexion with measles. He calculated at two periods —1891–1902 and 1903–12—the mean age of attack in a rural and in an urban community and found on both occasions that the former had the higher mean age.

Mortality and type of dwelling

That environmental conditions influence the age mortality in this manner is evident in the statistics of Glasgow for the years 1909–12. The population and deaths from specific diseases during that period were classified according to age and type of dwelling—whether one-, two-, three- or four-apartment houses. The final rates were published in the Medical Officer's *Annual Report* for the year 1912, and in Table IV we present in certain age periods the mortality from a group of infectious diseases.

It will be observed that when the mortality at age 0–1 year in the one-apartment house is represented by 100, that in the four-apartment house is 36: at age 1–5 the difference is still more outstanding as the death rate in the best type of house is only 17 % of that in the presumably most overcrowded dwelling.

TABLE IV

The male death rate from a group of infectious diseases per* 1000 *of the population according to the type of house in Glasgow*, 1909–12

Type of houses	0–1	%	1–5	%	5–15	%
One-apartment houses	49·14	100	19·19	100	1·96	100
Two-apartment houses	38·55	78	13·02	68	1·84	94
Three-apartment houses	20·95	43	7·78	41	1·47	75
Four-apartment houses	17·63	36	3·19	17	2·52	129

* Diphtheria, scarlet fever, measles, whooping cough, diarrhoea and enteritis.

In the next age period, 5–15, the sequence is completely changed as the mortality of well-housed children is now 29 % higher than that in the one-apartment house. Why has the trend of the mortality in this social range differed as between pre-school and school age? There is only one adequate explanation. The children in the worst environment—the one-apartment house—in addition to being possibly of a lower nutritional standard, had been infected in the pre-school life, age 0–5, with a resultant high mortality, whereas the children in the highest social class were not seriously exposed to infection until they attended school.

OVERCROWDING AND MORTALITY IN LONDON BOROUGHS

The effects of environment on health can be suitably studied in the boroughs of London. If we accept overcrowding, i.e. the percentage of the population living more than two in a room, with its physical, mental and economic implications, as an indication of an unhealthy environment, then these areas represent a wide range of hygienic conditions. The range of overcrowding at the 1931 census was from 4 % in Hampstead to 29 % in Shoreditch and Finsbury. To assess the extent to which the mortality is associated with environment we have correlated the death rates at age periods amongst females in each of the boroughs with the overcrowding index. We specifically selected females because they, and certainly the mothers, are more exposed to risks of their particular environment than are the males, who in all probability work outside it.

The trend of the coefficients in Table V clearly indicates the necessity of taking age into consideration in any discussion on the influence of environment on health. The coefficient at age 0–1 is $0{\cdot}405 \pm 0{\cdot}158$ but, in the next age group, it is no less than $0{\cdot}813 \pm 0{\cdot}064$. This latter value is the second largest coefficient in the series, and it confirms our previous discovery that, in childhood, the age 1–2 years is now the most responsive to hygienic conditions. After this age the coefficient becomes progressively smaller and is at a minimum, $0{\cdot}334 \pm 0{\cdot}168$, at age 4–5 years. The upward trend begins again, but it is slow at first. It becomes more defined as middle age is reached and culminates in the highest

TABLE V

The correlation coefficients between the female mortality from all causes and overcrowding in the London boroughs, 1929–33

Ages	r and S.E.
0– 1	0·405 ± 0·158
1– 2	0·813 ± 0·064
2– 3	0·522 ± 0·137
3– 4	0·396 ± 0·159
4– 5	0·334 ± 0·168
5–15	0·356 ± 0·165
15–25	0·362 ± 0·164
25–45	0·518 ± 0·138
45–65	0·910 ± 0·032
65–75	0·794 ± 0·069
75+	0·650 ± 0·109

peak value of 0·910 ± 0·032 at age 45–65. We thus see that environmental influence on mortality is strongest at two periods of life, at age 1–2 and at age 45–65. Its occurrence at these ages as revealed by the statistics for London is not a mere chance happening. It is also characteristic of the statistics of other places. We have previously seen it demonstrated at the younger age by the ratios of the mortality of the county boroughs in Northumberland and Durham to that of the rural districts in Eastern England while, at the older age, Brownlee, as will be seen below, using a narrower age limit than 45–65, indicated its existence in the age group 45–50, when he expressed the death rate at this age in Salford and in E 6 (England and Wales), respectively, as a ratio of that in H 3 (Healthy Districts).

Age	Death rate		Ratios			
	H3 (1891–1900)		E6 (1891–1900)		Salford (1891–1900)	
	M	F	M	F	M	F
35–40	6·26	5·76	1·44	1·36	1·59	1·90
40–45	7·57	6·82	1·58	1·47	2·51	2·14
45–50	9·32	7·83	1·60	1·50	2·67	2·41
50–55	12·55	10·22	1·56	1·47	2·42	2·39

The smallness of the correlation coefficients between the ages of 20 and 35 years is rather surprising in view of the fact that at this period of life tuberculosis is the most important cause of death and its incidence is higher in the slums than in residential districts. Hence it may well be asked: why should the

correlation between mortality and bad social conditions be more manifest at ages 1–2 and 45–65 years than at any others and what specific diseases were unduly affected? As far as the younger age is concerned we have previously incriminated the infectious group. At the older age period we are probably not witnessing any intensification in the effects of environment on the individual, but rather the result of accumulated strain of having long endured conditions of living which were deleterious to health. The strain would inevitably be most manifested in middle life—the period when the physiological mechanism of women is most disturbed. To obtain some idea of the diseases responsible we abstracted the important specific causes of death at this age, 45–65, and correlated their death rates in the various boroughs with the corresponding overcrowding values. The results were as follows:

	Values of r and S.E.
Respiratory tuberculosis and overcrowding	$0{\cdot}688 \pm 0{\cdot}100$
Other respiratory diseases and overcrowding	$0{\cdot}837 \pm 0{\cdot}056$
Cancer and overcrowding	$0{\cdot}357 \pm 0{\cdot}153$
Circulatory diseases and overcrowding	$0{\cdot}803 \pm 0{\cdot}067$
Other diseases and overcrowding	$0{\cdot}731 \pm 0{\cdot}088$

The correlation with cancer is small, but it is high with the other diseases and suggests that bad hygienic conditions are associated with general ill-health rather than with one specific cause.

Before concluding the mortality aspect of our investigation one particular point needs some explanation. We have previously declared that the relationship between density and mortality is best described by a curvilinear equation

$$m = c\delta^n,$$

and our subsequent adoption of the correlation coefficient which implies linearity seems rather illogical. It really is not so. Hitherto we were solely concerned with describing the association between density and the total mortality *at all ages*. But even if the relationship was non-linear at each specific age, we could deduce from the slopes of the best fitting straight lines the age period in which the association between the two variables was the most defined.

MORBIDITY FROM INFECTIOUS DISEASE AND ENVIRONMENT

The complete effects of environment on health are inadequately expressed when measured in terms of mortality because there may be a considerable amount of sickness in a population, yet the patients may not die. There are many diseases for which the number of deaths or the death rate is no criterion of the general prevalence. Scarlet fever is a classic example. The incidence of that disease—the number of notified cases as a ratio of the number of the population exposed to risk—at age 0–15 years is practically as high now as it was thirty or forty years ago, but the killing power of the disease is not nearly

so intense. We thus have a picture of a low mortality accompanying a high morbidity. Even the adoption of case rates as an accurate index of prevalence is in a sense insufficient, because we cannot be certain of either complete notification or correctness in diagnosis of the notified cases. In London, the diagnostic error for scarlet fever is approximately 10 % of the cases admitted to hospital. Roughly a quarter of the cases sent to hospital as diphtheria are suffering from something else—tonsillitis or laryngitis; while for enteric the error is in the neighbourhood of 30 %. But despite the inaccuracy with which the case rates of infectious disease are invested, they nevertheless convey a more complete picture of the presence of infection in the general population than is possible from a study of the mortality. Hence it is reasonable to suppose that the influence of environment would be more clearly demonstrated with morbidity than with mortality, assuming that adverse environmental conditions, as expressible in terms of overcrowding, are deleterious to health.

Nowhere, it would seem, can the relationship between morbidity and unhealthy conditions of living be better examined than in the boroughs or sanitary divisions of a large city, because these areas possess a homogeneity which is not so apparent in the different sections of the whole community. If our supposition is correct, and *a priori* it seems reasonable, that bad environmental conditions are inimical to health, then we should expect to find fairly high positive correlation between the variables in question. The association cannot be perfect, because where there is a high concentration of density there will inevitably be some degree of immunity against infection acquired by sub-clinical attack.

Although we have suggested that the relationship between density and disease is best measured within the sanitary or administrative subdivisions of a city it does not follow that the association will be equally, or even approximately, the same for different cities or for different infections in the same city. The correlation between overcrowding and the case rate at age 0–15 years for scarlet fever in Glasgow and in London supports this viewpoint:

Scarlet fever and overcrowding

Period	London	Period	Glasgow
1901–10	$r = +0{\cdot}135 \pm 0{\cdot}186$	1899–1902	$r = -0{\cdot}860 \pm 0{\cdot}045$
1911–14	$r = -0{\cdot}353 \pm 0{\cdot}165$	1903–08	$r = -0{\cdot}861 \pm 0{\cdot}054$
1919–23	$r = -0{\cdot}064 \pm 0{\cdot}188$	1909–13	$r = -0{\cdot}663 \pm 0{\cdot}120$

In London the prevalence of scarlet fever is little influenced by social conditions as the coefficients are very small. If there be a relationship it is of a slightly inverse character, as in two out of three instances the coefficients are negative. On the other hand, in Glasgow there is a well-marked tendency for residential districts to have relatively more scarlet fever than the poorer areas.

A plausible explanation of this phenomenon in Glasgow may be the greater immunization by minimum dosage in the slums than in the better class districts. But why the differentiation between the two cities? Possibly the type of housing in Glasgow is the responsible factor. There is a great difference between the housing systems in the two cities, as is apparent from the following facts obtained at the 1921 Census.

No. of rooms (a)	Percentage of population				Rooms per person	
	London (b)		Glasgow (c)		London (d)	Glasgow (e)
1	6·2	80·2	13·2	94·8	0·55	0·31
2	17·5		51·5		0·64	0·42
3	23·8		20·8		0·78	0·63
4	21·2		6·3		0·90	0·85
5	11·5		3·0		1·04	1·15

The distribution of the population in London according to the number of rooms occupied is fairly symmetrical, whereas in Glasgow the curve of incidence is rather skew. We find that 80·2 % of the total population of London and 94·8 % of that of Glasgow lived in homes containing one to five rooms. The disproportion was more strongly marked at the bottom end of the scale. In London, 6·2 and 17·5 % of the population lived in homes of one and two rooms: the corresponding proportions in Glasgow were 13·2 and 51·5 %.

These figures are, in themselves, not necessarily indicative of overcrowding, because the smaller proportions of the population in London—6·2 and 17·5 %—could be composed of families containing one or more members, whereas the constitution in Glasgow could be that of individual members. (According to Census regulations, a lodger occupying part of a house or flat is treated as a separate family.) But when the data in cols. (*b*) and (*c*) are supplemented by those in cols. (*d*) and (*e*) they demonstrate clearly the unsatisfactory position of Glasgow. The range between bad and good conditions is more accentuated in Glasgow than in London; the room space per person extends from 0·31 in the one-room house to 1·15 in the five-room house in Glasgow, the comparable values in London being 0·56 and 1·04. Amongst the sections of the total population living in one and two rooms in Glasgow the room space per person was 44 and 35 % respectively less than in London. Arising out of this greater concentration or massing of the population in tenement dwellings with deficient room space per person there will inevitably be greater opportunity in Glasgow than in London of acquiring immunity to the disease.

On the other hand it may be suggested that the differences between the correlation values for the two cities, and especially the large negative coefficient

for Glasgow, may be due to variation in the standard of notification in the two cities. If there exists for the Glasgow slum children incomplete notification of the disease in the pre-school stage, more complete notification at the school age, and, in the residential areas, complete notification for all children, then, on this hypothesis, there will be a negative correlation between incidence and overcrowding. Brownlee was of the opinion that scarlet fever in Glasgow was a milk-borne infection, and as children in the residential areas consumed relatively more milk than the children in the poorer districts the higher incidence of scarlet fever in the former followed as a consequence. Hence we have three possible explanations and there may be others, but in the light of the abnormal type of housing in Glasgow we are inclined to adhere to the opinion that the phenomenon is best explained by the greater degree of latent immunity amongst those children who live under bad environmental conditions. This view is reinforced by the fact that the diphtherial experience in Glasgow and in London respectively is practically identical with that for scarlet fever.

Although the type of environment is inversely related to the incidence of infections such as scarlet fever and diphtheria, yet when it is correlated with the fatality arising from those infections the association is highly positive, as will be seen from the following values for Glasgow:

Period	r and S.E.
1899–1902	$0{\cdot}586 \pm 0{\cdot}114$
1903–08	$0{\cdot}789 \pm 0{\cdot}079$
1909–13	$0{\cdot}727 \pm 0{\cdot}096$
1921–25	$0{\cdot}619 \pm 0{\cdot}092$

There is nothing abstruse in the interpretation of these values. The slum child in an overcrowded home with implications of earlier age of infection and malnutrition is less able than the child in the higher social grades to resist a fatal attack.

There are other notifiable diseases the contraction of which as far as is known confers no subsequent absolute immunity, but which are definitely related to prevailing hygienic conditions. Erysipelas exemplifies this class and its statistical experience in Glasgow and in London at different periods is as follows:

Erysipelas and overcrowding

Period	London	Period	Glasgow
1901–10	$r = +0{\cdot}834 \pm 0{\cdot}058$	1899–1902	$r = +0{\cdot}515 \pm 0{\cdot}128$
1911–14	$r = +0{\cdot}641 \pm 0{\cdot}111$	1903–08	$r = +0{\cdot}698 \pm 0{\cdot}105$
1919–23	$r = +0{\cdot}745 \pm 0{\cdot}084$	1909–13	$r = +0{\cdot}713 \pm 0{\cdot}103$

The high positive correlation is probably due to the fact that in congested areas there is greater likelihood of abrasions being followed by a supervening infection with the streptococcus erysipelatis.

Our next aim was the presentation of a general picture of the relationship between the morbidity from infectious disease and the range of environmental conditions in the different administrative areas of England and Wales. Accordingly we abstracted the recorded number of notified cases of scarlet fever, diphtheria, enteric and erysipelas for the period 1921–30 in the London boroughs, each county borough and each urban and rural district and correlated the case rates with the mean value of the corresponding overcrowing index as recorded at the 1921 and 1931 censuses. The coefficients obtained for the different areas according to their geographical location are given in Table VI.

We are mindful of the fact that the values obtained are influenced by one important consideration—we were unable to make any allowance for variation in the standard of notification in the different areas. This factor may not be of serious import within the administrative areas such as the county boroughs because the percentage of all cases notified may not vary very much from city to city. In all probability it will affect comparison made between the administrative areas such as county boroughs and rural areas, as it is most unlikely that the standard of notification is the same in town and country. Many of the values in the table are statistically unimportant, but there are points of interest —the chief of which is the relationship between the incidence of erysipelas and environment. There is a positive correlation between the two variables in every part of the country but it diminishes with decreasing urbanization. In London the coefficient was $0{\cdot}665 \pm 0{\cdot}103$: in the aggregate rural districts $0{\cdot}330 \pm 0{\cdot}039$. The incidence of scarlet fever in London is positively but insignificantly related to environment—as in past experience—but that of diphtheria is much more definite, r being $0{\cdot}363 \pm 0{\cdot}161$. It will be noted that the statistical experience of these two diseases in the rural districts resembles that of London inasmuch as the association is positive. The relationship exhibited in the county boroughs, particularly in the North, is entirely different from that elsewhere. The association is negative, resembling in this respect that of Glasgow. The concordance is not surprising, as the environmental conditions of the northern towns, in all probability, are as unsatisfactory as those of Glasgow and the opportunity of acquiring immunity to either disease by a sub-clinical attack is probably just as readily obtained.

The association with enteric is rather vague as, apart from that for the urban districts in Wales, not one of the coefficients is statistically important and the signs are almost evenly distributed. In the component regional divisions the negative sign occurs six times and the positive five times. In view of the sporadic occurrence of this disease and the different channels by which the infection may be conveyed, particularly by "carrier", the indefinite relationship

TABLE VI

Correlation coefficients based on the statistics for 1911–20

Area (excluding all districts with population less than 5000 persons)	No. of observations	Correlation coefficient r between the percentage of population living more than two in a room and the following case rates:			
		Scarlet fever case rate per 1000 population aged 0–15 years	Diphtheria case rate per 1000 population aged 0–15 years	Enteric fever case rate per 1000 population all ages	Erysipelas case rate per 1000 population all ages
		r and S.E.	r and S.E.	r and S.E.	r and S.E.
Metropolitan Boroughs					
London	29	$0{\cdot}1263 \pm 0{\cdot}183$	$0{\cdot}3628 \pm 0{\cdot}161$	$-0{\cdot}1160 \pm 0{\cdot}183$	$0{\cdot}6654 \pm 0{\cdot}103$
County Boroughs					
North	41	$-0{\cdot}4068 \pm 0{\cdot}130$	$-0{\cdot}3440 \pm 0{\cdot}138$	$0{\cdot}1501 \pm 0{\cdot}153$	$0{\cdot}2353 \pm 0{\cdot}148$
Midlands, South and Wales	41	$-0{\cdot}0561 \pm 0{\cdot}156$	$0{\cdot}1127 \pm 0{\cdot}154$	$-0{\cdot}2260 \pm 0{\cdot}148$	$0{\cdot}2106 \pm 0{\cdot}149$
England and Wales	82	$-0{\cdot}1062 \pm 0{\cdot}109$	$-0{\cdot}1626 \pm 0{\cdot}108$	$-0{\cdot}0638 \pm 0{\cdot}110$	$0{\cdot}4034 \pm 0{\cdot}092$
Urban Districts*					
North	241	$-0{\cdot}1560 \pm 0{\cdot}063$	$-0{\cdot}0850 \pm 0{\cdot}064$	$0{\cdot}0139 \pm 0{\cdot}064$	$0{\cdot}2220 \pm 0{\cdot}061$
Midlands	210	$-0{\cdot}0496 \pm 0{\cdot}069$	$0{\cdot}1455 \pm 0{\cdot}068$	$-0{\cdot}1832 \pm 0{\cdot}067$	$0{\cdot}1634 \pm 0{\cdot}067$
South	145	$0{\cdot}1204 \pm 0{\cdot}082$	$0{\cdot}2887 \pm 0{\cdot}076$	$-0{\cdot}0001 \pm 0{\cdot}083$	$0{\cdot}1354 \pm 0{\cdot}082$
Wales	57	$-0{\cdot}1256 \pm 0{\cdot}130$	$-0{\cdot}3644 \pm 0{\cdot}115$	$0{\cdot}3535 \pm 0{\cdot}116$	$0{\cdot}1606 \pm 0{\cdot}129$
England and Wales	653	$0{\cdot}0281 \pm 0{\cdot}039$	$-0{\cdot}0866 \pm 0{\cdot}039$	$0{\cdot}0170 \pm 0{\cdot}039$	$0{\cdot}3439 \pm 0{\cdot}035$
Rural Districts					
North	110	$0{\cdot}1605 \pm 0{\cdot}093$	$0{\cdot}0215 \pm 0{\cdot}095$	$0{\cdot}0227 \pm 0{\cdot}095$	$0{\cdot}3864 \pm 0{\cdot}081$
Midlands	214	$0{\cdot}1569 \pm 0{\cdot}067$	$0{\cdot}2989 \pm 0{\cdot}062$	$-0{\cdot}0296 \pm 0{\cdot}068$	$0{\cdot}2225 \pm 0{\cdot}065$
South	152	$0{\cdot}2706 \pm 0{\cdot}075$	$0{\cdot}2360 \pm 0{\cdot}077$	$0{\cdot}0113 \pm 0{\cdot}081$	$0{\cdot}0734 \pm 0{\cdot}081$
Wales	49	$0{\cdot}2501 \pm 0{\cdot}134$	$0{\cdot}2336 \pm 0{\cdot}135$	$-0{\cdot}0008 \pm 0{\cdot}143$	$0{\cdot}0623 \pm 0{\cdot}142$
England and Wales	525	$0{\cdot}2106 \pm 0{\cdot}042$	$0{\cdot}1114 \pm 0{\cdot}043$	$-0{\cdot}0228 \pm 0{\cdot}044$	$0{\cdot}3302 \pm 0{\cdot}039$

* Urban districts include municipal boroughs.

between the variables as exhibited by the series of coefficients in the table is, in a sense, not surprising.

Conclusions

The points of interest and the conclusions warrantable from this study are:

1. The relationship between general mortality and density of population or overcrowding is appropriately described by an equation of the character used by Farr:

$$\text{Mortality} = \text{C. Density}^{x}.$$

For a certain range of density there is a corresponding increase in the mortality, but a saturation level is reached when for further increase in density there is no accompanying increase in mortality. The statistics of overcrowding in the London boroughs when plotted against the corresponding standardized death rates for the period 1930–2 reveal a definite curvilinear relationship (Diagram on p. 346).

2. The effects of a bad environment on health are particularly noticeable at two periods of life—in childhood and in middle life.

Accepting the mortality in town and in country, respectively, as representing indices of two widely different environments it was found that during childhood the greatest divergence occurred at age 2–3 years. This was a characteristic feature until 1911. Afterwards there was a transition and age 1–2 years takes priority (Tables I and II). This fact is further confirmed by the correlation coefficients between overcrowding and age mortality in London, as the highest coefficients were those at age 1–2 years and at age 45–65 years (Table V).

The diseases responsible for the initial manifestation (the high ratio at 2–3 years) were mainly those in the infectious group (Table III). At the older age period it would appear that bad hygienic conditions are associated with general ill-health rather than with one specific cause.

3. Type of dwelling is highly related to the mortality from infectious disease, as is observable in the statistics for Glasgow (Table IV). Children living in single-room dwellings have a very high mortality in pre-school life as compared with children living in larger sized homes. A probable explanation is that they are sooner exposed to infection and less able to withstand a fatal attack. Children in the better class house get infection when they come to school, as is evident from their higher mortality at age 5–15 years.

4. Morbidity is a better index of environmental influence than mortality because, for diseases such as scarlet fever, mortality is no criterion of its prevalence. The incidence of scarlet fever and diphtheria is negatively correlated with overcrowding in Glasgow but not in London. The difference or distinction may be due to a possibly greater degree of latent immunity amongst Glasgow children as a consequence of the unique housing conditions in that city.

5. The response of infection to environment differs considerably both for type and location of administrative area (Table VI). Undoubtedly some of the

difference between the coefficients for town and country areas is due to varying standards in the notification of infectious disease. The relationship between overcrowding and the incidence of scarlet fever and diphtheria in the county boroughs of the North is an inverse one—similar to that in Glasgow and probably capable of a similar explanation.

Erysipelas is an instance of a particular disease definitely associated with hygienic conditions, as the correlation coefficients are positive in all parts of the country. The correlation between environment and enteric is very indefinite, but this is probably due to the many factors which can be responsible for the spread of this particular infection.

REFERENCES

BROWNLEE, J. (1922). "The use of death rates as a measure of hygienic conditions." *Spec. Rep. Ser. Med. Res. Coun., Lond.*, No. 60.

FARR, W. (1843). *Registrar-General's Fifth Annual Report*, Appendix, pp. 420–4.

FARR, W. (1875). *Supplement to Registrar-General's 35th Annual Report*, 1861–70.

PICKEN, R. M. F. (1921). "Epidemiology of measles in a rural and residential area." *Lancet*, June, 1921.

ROYAL STATISTICAL SOCIETY DISCUSSION (1922). *J.R. Statist. Soc.* **85**, 537.

TATHAM, J. (1895). *Supplement to Registrar-General's 55th Annual Report*, 1881–90.

ON SENTENCE-LENGTH AS A STATISTICAL CHARACTERISTIC OF STYLE IN PROSE: WITH APPLICATION TO TWO CASES OF DISPUTED AUTHORSHIP

By G. UDNY YULE

Section I. Introductory

One element of style which seems to be characteristic of an author, in so far as can be judged from general impressions, is the length of his sentences. *This* author develops his thought in long, complex and wandering periods: *that* finds sufficient for his purpose a sequence of sentences that are brief, clear and perspicuous. Since the length of a sentence can be readily measured, for practical purposes, by the number of words, it occurred to me that it would be of interest to subject this impression to statistical investigation.

In carrying out the investigation, I met with more difficulties than I had foreseen. There are two terms used above: (1) Sentence, (2) Word. What is a sentence? What is a word, or what for present purposes is to be regarded as a word?

Sentence. Let me cite the *New English Dictionary*:

> Sentence. *sb.* 6. A series of words in connected speech or writing, forming the grammatically complete expression of a single thought; in popular use often (= Period *sb.* 10) such a portion of a composition or utterance as extends from one full stop to another. In *Grammar*, the verbal expression of a proposition, question, command, or request, containing normally a subject and a predicate (though either of these may be omitted by ellipsis). In grammatical use, though not in popular language, a sentence may consist of a single word....English grammarians usually recognise three classes: simple sentences, complex sentences (which contain one or more subordinate clauses), and compound sentences (which have more than one subject or predicate).

From these definitions I conclude, I hope rightly, that we may drop the term "period" and use the term "sentence" to cover *any* sentence (or as I should have been inclined to write "period"), however complex and however compound in the senses defined. It is convenient to be able to avoid a term which to a statistician would generally suggest a different meaning. Now, not being a grammarian but just one of the populace, I confess that I started with the popular notion of a "sentence" in this general sense: "such a portion of a composition as extends from one full stop to another", and thought I would have nothing to do but tot up the words from full stop to full stop. The first definition, however, reads: "the grammatically complete expression of a single thought." I feel some doubts as to the "*single* thought". (Is not "I am tired and hungry" a sentence, and does it not convey two thoughts, the thought of being tired and the thought of being hungry?) But the "grammatically complete

expression" surely is essential to make a word-series a sentence; the word-series must be what Webster calls a "sense unit", and the trouble is that, especially in older works, "a portion of a composition" which "extends from one full stop to another" is often *not* the grammatically complete expression of anything. When the author or compositor has used punctuation in this fashion it is no longer possible simply to add up words from one full stop to the next, paying little or no attention to sense: it is necessary for the reader frequently to pull up and ask himself if the words just read do or do not form a sentence, and if they do not, what are in fact the limits of the sentence within which they must be assumed to lie. I need hardly point out how much this increases labour, and even, if the sentences are very long and complicated, brings in largely the element of personal judgement. Two readers, at least unskilled readers like myself, may well differ as to where a given sentence terminates.

Here is quite a simple illustration of the difficulty from a modern essay on *The Politics of Burns* (ref. 1, at end of paper):

> There are several points here all at once calling for notice, and seldom getting it from friends of the poet:
>
> The extraordinary talent for history shown by Robert Burns.
>
> His attention to British History in preference to Scottish.
>
> The originality of his views.

In this passage there are four word-series, the first divided from the second only by a colon (though the second begins with a capital letter), the second divided from the third, and the third from the fourth, by full stops. But neither the second, nor the third, nor the fourth word-series is a grammatically complete expression. The whole passage must be taken together, as it seems to me, as one single sentence. I am of course simply illustrating my difficulty, not criticizing the punctuation.

On the other hand, where an author has written a very long and meandering sentence, a question may well arise between two different readers as to whether a halt should not be called in the middle, and a full stop entered where author or compositor has placed only a colon.

I say author *or* compositor, for it must not be assumed that one is necessarily laying sacrilegious hands on the deliberate construction of the author himself. "So far as punctuation is concerned," says McKerrow (ref. 2), "there seems very little evidence that many authors exercised any care about it whatever. After all, even at present, few authors trouble to punctuate their MSS. with any care or consistency. Such punctuation as is found in ordinary MSS. of the sixteenth and seventeenth centuries is indeed most erratic and seldom goes beyond full stops at the end of most of the sentences and some indication of the caesura in verse." I had, before I started the present work, expected that this comment would apply much more to intermediate punctuation than to full stops, trusting that authors would at least insert "full stops at the end of *most* of their sentences".

But that it applies to both was enforced on me by different versions of the short tract by Gerson, *De Meditatione Cordis*, in the edition of his complete works that I used (see below section III and ref. 9) and in four editions of the *Imitatio Christi* on my shelves. The versions differed, not only verbally, but also as regards full stops. If punctuation, even as regards full stops, is largely the work of the compositor, there need be no hesitation in overriding them if necessary: indeed, the use of personal judgement seems unavoidable.

Let me add that at first I by no means realized the full extent of this difficulty, and when I did often felt myself horribly incompetent to deal with it. I am sure my final decisions could often be contested, and were not infrequently inconsistent with one another. But after all difficult cases are but a small proportion of all sentences in most writers and, if only as an exploratory piece of work, I hope the investigation may still retain interest and value.

Word. Compared with the difficulties as to the sentence, the difficulties concerning words are really of a minor kind. One large class is indicated by the lines of Calverley:

> Forever; 'tis a single word!
> Our rude forefathers deemed it two:
> Can you imagine so absurd
> A view?

Our rude forefathers also wrote *it self*, *any where*, *every where* and so forth, where their rude descendants write *itself*, *anywhere*, *everywhere*. How shall we reckon such expressions? It is best, I think, to follow modern usage and I generally endeavoured to do so; but in rapid counting it is very easy to make a slip. Hyphened words present the same sort of difficulty. *Law-courts*, *china-manufacturer*, *news-journal*, *well-earned*, I would count as two words each; *out-of-the-way* as four: but *co-acervation*, *contra-distinguish*, *tri-syllabic*, *pre-disposed*, *re-produce*, as one each. A *something-nothing-every-thing* (Coleridge) presents a special problem: I think it should be three words. But how many words is *matter-of-factness*? Coleridge calls it *a* word, "an uncouth and new coined word".

Then there are abbreviations such as *viz.*, *i.e.*, *etc.* or *&c.* The first there is no reason to reckon as anything but one word. The second, third and fourth, in spite of their meaning, I also reckoned as one each: eye and mind grasp them as wholes.

Finally, what are we to do with figures? Dates may occur even in literary or historical essays: any year stated in figures (1825 or 1798) I reckoned as a word. Whether days of the month ever occurred I do not recall: but I would reckon the day of the month stated in figures, as in January 10th, as a word for the month and a word for the number of the day. Any actual number if stated in figures, and such numbers are frequent of course in the work of Graunt and Petty that I have discussed, would be reckoned as one word whatever the

number. Thus 251 would be reckoned as one word and so would 3,251,452; although two hundred and fifty one would be five words, and three million, two hundred and fifty one thousand, four hundred and fifty two would be thirteen. This may seem arbitrary: but again, if the number is stated in figures eye and mind grasp it as a whole, while if in words it has to be taken word by word. For the same reason, fractions such as $\frac{3}{4}$ or $\frac{1}{2}$, which are also frequent in Graunt and Petty, were reckoned as a word each. Sums of money stated in figures, such as £1. 2*s*. 8*d*. were to the best of my recollection treated as if pounds, shillings and pence were so expressed in words—not very consistently with the principle stated above. If any matter was so full of figures that it practically ceased to be prose even in the humblest sense of that term, if for example it was set out in tabular or semi-tabular form, it was simply cut out.

In all such instances as the above I really do not think it is of very much practical consequence what rule is adopted: nor even of much practical consequence if the treatment is not always self-consistent. Sentences vary too much in length for what are after all minor errors of measurement to be of much consequence.

Quotations. I may mention in conclusion one other difficulty. What is to be done with quotations? Two cases seem clear. If the author makes a brief quotation forming grammatically part of his own sentence, he is only substituting someone else's words for his own and they must be counted in: as in Lamb's

> But I am none of those who—
> *Welcome the coming, speed the parting guest.*

If, on the other hand, the author simply quotes a complete sentence from somebody else, *that* is not the author's writing and must be omitted: as for example when the same author writes

> A *gag-eater* in our time was equivalent to a *goul*, and held in equal detestation. — suffered under the imputation.
> —'*Twas said*
> *He ate strange flesh.*

The quotation must be dropped. But no rule can be applied strictly to living literature. Thomas à Kempis, for example, quotes the words of scripture so freely that if one cut out scriptural quotations one would eliminate a considerable proportion of his work. He has made scripture his own, and what he has written must stand as his.

A serious difficulty arises only when, say, an essayist is discussing a poet and makes a long and purely illustrative quotation. This may be of any length, and it may be so made as virtually to form part of the sentence of the critic himself, or may follow almost indifferently a colon or a full stop at the end of the critic's sentence. Quotations made in the first way, and even those made in the second way after a colon, I tended at first to include. But, on coming across *very* long

quotations, it became obvious that this was unsatisfactory, and I then adopted the easier method of simply cutting out all pages on which this source of trouble was serious. This is, I think, the best course.

SECTION II. ILLUSTRATIONS FROM BACON, COLERIDGE, LAMB AND MACAULAY

This section is in part purely illustrative, showing what sort of distributions of sentence-length we may expect, but in part is concerned with the fundamental question, how far sentence-length is really a *characteristic* of an author's style. If, that is to say, we take two lengthy passages, each containing a few hundred sentences, from a given fairly homogeneous work, will they present us with proportional numbers of sentences of each particular length in reasonably close agreement with one another? If they do not; if, although dealing with the same sort of material in the same sort of way, the author is liable capriciously to vary in the length of his sentences, sentence-length is not a *characteristic* of his style in any proper sense of the term, and one's impression to the contrary will be proved mistaken. If, however, there is reasonably close agreement, we can accept sentence-length as a characteristic. It is necessary, I think, to insert the condition that the author shall be dealing with the same sort of material in the same sort of way, since (again judging from general impressions) it seems clear that sentence-length may be affected by the author's matter as well as by his individuality: argumentative passages, for example, may well tend to longer sentences than matter purely descriptive.*

The four authors chosen as illustrations are Bacon, Coleridge, Lamb and Macaulay; and their works, Bacon's *Essays*, Coleridge's *Biographia Literaria*, Lamb's *Elia* and *Last Essays of Elia*, and Macaulay's *Essays*. The particular editions used are not probably of any importance in this instance but are cited in the references at the end of the paper. They were simply those that I happened to have on my shelves.

The fundamental tables, all in the same form and showing the numbers of sentences with 1 to 5, 6 to 10, 11 to 15 words, and so on, are given in the Appendix.

Table A gives the data derived from Bacon's *Essays*. Here, when I had got to the end of Essay XXVI, "Of Seeming Wise", I judged myself to be about half-way, and called this batch of 462 sentences sample A: I then proceeded to the end of Essay LI, "Of Faction", and as this had given me 474 sentences, or approximately the same number, I called it sample B. The total number of essays being 58, the two samples together cover almost 90% of the essays. Table A shows, in addition to the distributions for the two samples

* Compare, for example, in Hazlitt's *Lectures on the English Comic Writers*, the style of the first essay "On Wit and Humour" with that of the subsequent lectures on definite groups of writers. See also below, section IV, for some remarks on Petty.

A and B, the total distribution for the two together. From inspection it will be clear that the two samples are very concordant, though figures are inevitably slightly irregular and fluctuating. In both the frequencies increase rather abruptly in the interval 11–15; in both they reach a maximum in the interval 31–35, and then tail away very slowly indeed, so that there is a considerable number of sentences of 101–200 words in length and a few over 200. The record is a sentence of 311 words, as punctuated, i.e. from full stop to full stop. The reader will find it in the penultimate paragraph of Essay XXVII, "Of Friendship". It might well be broken up: but I do not think at this early stage I had attempted any revision of punctuation, hardly having realized the difficulty mentioned in the preceding section.

Table B gives the data from Coleridge's *Biographia Literaria*. I began at the beginning and continued to about the middle of chapter IX, when I had a batch of just over 600 (actually 601) sentences, which I judged sufficient. this is sample A. For sample B I meant to take a similar batch from near the end and began with chapter XX in vol. II, not noticing that a great part of the remainder of this volume consisted of "Satyrane's Letters". The result was that chapter XX to the end gave me only about half the number of sentences wanted, and to complete the sample I went back to the beginning of the volume (chapter XIV) and worked on from that point to about the middle of chapter XVIII. This gave me sample B of 606 sentences. Again, inspection of the table shows that the distributions for samples A and B are closely alike and somewhat different from those of Table A. The actual maximum frequency occurs earlier, at 26–30 for sample A, and 21–25 both for sample B and for the two samples together; and the distribution is less scattered, there being a smaller proportion of the very long sentences of over 100 words in length. With *Biographia Literaria* the quotation difficulty became at times acute: a page or two, or a shorter passage, was omitted here and there to evade it.

The data derived from Lamb's essays are given in Table C. Sample A was taken from *Elia* (1st edition, 1823), from the beginning to some two-thirds of the way through "Mrs Battle's Opinions on Whist". Sample B was drawn from the middle of the *Last Essays of Elia* (1st edition, 1833), starting with the essay "Detached Thoughts on Books" and continuing to the end of "Barbara S—". Once more, the general consistence of the two samples looks quite satisfactory. Short sentences are much more frequent than with Coleridge, and the greatest frequencies occur in the intervals 6–10 and 10–15, which are almost equally frequent.

Finally, in Table D we have the data from Macaulay's *Essays*. Sample A was taken from the beginning of the essay entitled "Lord Bacon" (1837): sample B from the beginning of the essay on the Earl of Chatham (1844). In this instance the two samples do not agree quite so well as in previous tables. The first three frequencies are quite concordant and agree in placing the maximum frequency

at sentences of 11–15 words. But thereafter the frequencies of sample B exceed those of sample A right up to the interval 46–50, after which the position is reversed, so that the second sample is less scattered than the first. But the difference is not great.

So far we have dealt only with the similarities and differences suggested by brief inspection of the tables, but it is desirable to summarize in terms of statistical measures. Distributions of this kind, with long tails in which rather wild outliers may occur, might, it seemed to me, be best dealt with by the method of percentiles. While therefore I have calculated the arithmetic means as the most familiar form of average, I have also given the median, and for the rest have contented myself with the lower and upper quartiles Q_1 and Q_3, the interquartile range $Q_3 - Q_1$ as a measure of dispersion, and the ninth decile D_9 as an index to the extension of the tail of the distribution. These percentiles are calculated on the usual convention that the intervals may be regarded as 0·5–5·5, 5·5–10·5, 10·5–15·5, etc., and the distribution treated as continuous.*

These constants, for Tables A–D, are given in Table I. The table brings out very well the degree of consistence of each author with himself, and his differences from the others. For samples A and B of Bacon, mean, median, lower quartile and interquartile range agree within less than a unit, upper quartiles differ by 1·5 units and ninth deciles by 2·4, no very great difference from the practical standpoint especially in the constants most affected by fluctuations of sampling. For Coleridge, the two samples differ by between 1 and 2 units in the case of mean, median and lower quartile; the upper quartiles differ by 3·3, the interquartile ranges by 2·1 and the ninth deciles by 4·2. For Lamb the differences are less than a unit in the case of mean, upper quartile and interquartile range, the difference is exactly a unit for the two lower quartiles, 1·3 units for the medians, and 3·6 units for the ninth deciles. For Macaulay the

* As offprints at least of this paper may fall into the hands of some who are not statisticians, I may be forgiven for a note of explanation. The arithmetic mean is the common form of average, the sum of the quantities to be averaged divided by their number. Given a frequency distribution, it is calculated on the assumption that all observations falling into any one interval have the mid-value of that interval, e.g. that all sentences in the interval 6–10 are eight words long: this gives quite a close approximation. The lower quartile is the sentence-length such that one quarter of all sentences are shorter and three quarters longer. But sentence-lengths are discontinuous: sentences of 25 words or less might be less than a quarter of the whole, sentences of 26 words or less more than a quarter; hence some convention is necessary if a precise value is to be stated. The convention is that given in text above, and we proceed by simple interpolation. Thus in the total distribution of Table A the total number of sentences is 936, one quarter of which is 234. The first four frequencies up to and including sentences of 25 words, or up to the conventional limit 25·5, give a total of 212, and accordingly we require 22 more. There are 85 in the next interval, which is an interval of five words, and the lower quartile is therefore approximately

$$25{\cdot}5 + \frac{22}{85} \times 5 = 26{\cdot}8.$$

The upper quartile, the value exceeded by only one-quarter of the observations, and the ninth decile, the value exceeded by only one-tenth, are similarly determined.

TABLE I

Constants for the distributions of sentence-length in samples from Bacon, Coleridge, Lamb and Macaulay (Tables A, B, C and D of Appendix). Q_1 = Lower Quartile, Q_3 = Upper Quartile, D_9 = Ninth Decile)

Constant	Bacon			Coleridge		
	A	B	Total	A	B	Total
Mean	48·4	48·5	48·5	41·2	39·5	40·3
Median	39·4	39·4	39·4	35·7	34·2	34·9
Q_1	27·2	26·4	26·8	22 9	21·8	22·3
Q_3	61·7	60·2	60·9	53·2	49·9	51·3
Q_3-Q_1	34·5	33·8	34·1	30·3	28·1	29·0
D_9	89·5	91·9	91·0	74·5	70·3	73·1
	Lamb			Macaulay		
	A	B	Total	A	B	Total
Mean	26·2	26 3	26·2	22·8	21·4	22·1
Median	18·3	19·6	19·1	18·2	18·9	18·6
Q_1	10·5	11·5	11·0	11·5	12·0	11·7
Q_3	33·3	34·0	33·7	28 2	27·5	27·8
Q_3-Q_1	22·8	22·5	22·7	16·7	15·5	16·1
D_9	57·5	53·9	54·9	44·2	39·1	40·6

constants seem almost more self-consistent than inspection of the table would lead one to expect. The differences are, for means 1·4, medians 0·7, lower quartiles 0·5, upper quartiles 0·7, interquartile ranges 1·2, ninth deciles 5·1: the lessening of the scatter has affected mainly the ninth decile. For Coleridge all the constants given are lower than the corresponding constants for Bacon, the differences being most conspicuous for the upper quartile and the ninth decile. Comparing Lamb and Macaulay, medians and lower quartiles are much the same, but Macaulay's mean, upper quartile, interquartile range and ninth decile are appreciably lower than the corresponding figures for Lamb.

We may conclude accordingly that sentence-length *is* a characteristic of an author's style. There is no discrepancy between the results of our statistical investigation and the judgement made from general impressions. Given similar material and mode of treatment, an author's frequency distribution of sentence-lengths does remain constant within fairly narrow limits. At the same time, it must be admitted, the limits cannot be precisely defined. In case of dispute as to whether two works are or are not by the same author, a judgement based on frequency distributions of sentence-lengths for the two must in the end be a

personal one, and founded on such differences as are observed between samples from works known to be by the same author. Hence the importance of the illustrations that have been given.

The test is numerical, but not exact. For there can be no question of applying the ordinary tests based on the theory of simple sampling. The "samples" we have taken are in no sense random samples: they are continuous passages, or collections of continuous passages, and if (as was my practice) the lengths of sentences are written down in order as they occur it is very clear that the resulting numerical series is not a random series but a "clumped" series. Short sentences tend to occur together. The tendency is much clearer for some authors than for others and for Macaulay is a characteristic trick of style, a point being emphasized by a series of hammer-blows from sentences of very few words: for example,

> These are the old friends who are never seen with new faces, who are the same in wealth and in poverty, in glory and in obscurity. With the dead there is no rivalry. In the dead there is no change. Plato is never sullen. Cervantes is never petulant. Demosthenes never comes unseasonably. Dante never stays too long.

Or again,

> The two sections of ambitious men who were struggling for power differed from each other on no important public question. Both belonged to the Established Church. Both professed boundless loyalty to the Queen. Both approved the war with Spain.

It is obvious that a series formed from the lengths of such sentences is not a random one and that consequently differences between samples taken as we have taken them may greatly exceed the limits of *simple sampling* without, for practical purposes, being of any real significance. The differences between the upper quartiles and between the ninth deciles of the two samples from Coleridge, for example, are 10 or 11 times the standard errors, but cannot be regarded as very material.

One point regarding the form of these distributions may be noted as of interest to the statistician. They are not of the Poisson type but of the type in which the square of the standard deviation largely exceeds the mean. The following are the figures for the total distributions, the unit being a word:

	M	σ^2	σ
Bacon	48·45	1048·22	32·38
Coleridge	40·34	677·10	26·02
Lamb	26·25	514·14	22 68
Macaulay	22·07	230·04	15·17

I now pass on to an application of the method to a case of disputed authorship

Section III. The authorship of the *De Imitatione Christi*: Thomas à Kempis and Gerson

Although the old controversy as to the authorship of the *Imitatio* still continues, and only last year a translation from Netherlandish texts was published in America (ref. 7) attributing it to Gerald Groote, the founder of the Brothers of the Common Life, few I believe will not hold it to have been definitely settled in favour of Thomas à Kempis. That certainly is my belief. Any reader who wants to know more of the evidence will find a brief summary in ref. 11, or a more detailed treatment in refs. 10, 12 and 13. If this does not suffice he can follow up De Backer's bibliography, ref. 14. But I thought it would be of some interest to see what results the present method would yield when applied to investigate the respective claims of Thomas à Kempis and one of those to whom the authorship was formerly attributed, Jean Charlier de Gerson, Chancellor of the University of Paris. That Gerson could have written the book seems plainly impossible since, apart from all questions of style, it was clearly written by one who was living the monastic life; but many early editions of the book bear his name, and in others the *Imitatio* is followed by Gerson's tractate *De Meditatione Cordis* almost as if it formed part of the same work.

Since many works of Thomas are extant, admitted as such even by those who deny his authorship of the *Imitatio*, we can deal with two problems: (1) does the distribution of sentence-length in the *Imitatio* resemble that in (other) admitted works by Thomas, or no?; (2) does the distribution of sentence-lengths in the *Imitatio* resemble that in the works of Gerson?

The edition of Thomas's works that I used was that of Pohl (ref. 8). In this edition the four books of the *Imitatio* are (to retain the usual numbering) placed, as in the Brussels autograph MS., in the order I, II, IV, III. The four books are of very different lengths, covering in this edition some 51, 29, 47 and 120 pages respectively. To get a sample fairly distributed over these books, in rough proportion to their respective lengths, I took ten subsamples of about 120 sentences each as follows: Lib. I, two, from the beginning and from near the end; Lib. II, one, from about the middle; Lib. IV, two, from the beginning and from near the end; Lib. III, five, distributed through the book. The subsamples from books I, II and IV together form sample A of Table E in the Appendix, and the five from book III, sample B. Sample B contains a rather higher proportion of very short sentences, but otherwise A and B are reasonably concordant. There was comparatively little trouble with the sentence-problem: Thomas was careful in punctuation, which may be taken as his own. But one point may be noted which occurs both in the *Imitatio*, in the miscellaneous works and in Gerson: it is a question arising from the punctuation of quotations or sayings. The following from the *Soliloquium Animae* will serve as an illustration:

Caeli dixerunt. Pertransivit nos et ascendit: invaluitque supra nos. Terra respondit. Si caeli caelorum non capiunt: nolite me interrogare. Stellae cecinerunt: tenebrae sumus et non lux si illuxerit. Mare contremuit et ait. Non est in me: et abyssus ignoravit.

Here there is a full stop after *dixerunt*, *respondit*, *ait*, before the words spoken are given, although after *cecinerunt* only a colon. In all cases, it seems to me, the words spoken or quoted should be counted in with the preceding words as if there was only a colon. Further, in Lib. III I have to confess to a piece of carelessness. A number of chapters in this book begin with the vocative "Fili" followed by a full stop. This should, I think, clearly be counted with the words following: in a translation it would be followed only by a comma. But at first I had entered the word as a one-word sentence, and did not realize that the point was important since this introduction was frequent. To have left things as they were would have created a misleading number of one-word sentences: to have revised the numbers of words in all the initial sentences of the chapters affected would have entailed more labour in altering tables than I was inclined to undertake. Finally, I simply struck out all these occurrences of initial "Fili", of which there were sixteen. Sentences in the *Imitatio* being very short, my original distributions were booked up ungrouped, and this made the number of "1's" very conspicuous.

The sample to represent the miscellaneous admitted works of Thomas à Kempis was similarly made up from ten subsamples of about 120 sentences each taken from the following:

(1) *De tribus Tabernaculis.*
(2) *Epistula ad quendam Cellerarium.*
(3, 4) *Soliloquium Animae.*
(5) *Meditatio de Incarnatione Christi.*
(6) *Sermones de Vita et Passione Domini.*
(7) *Hortulus Rosarum.*
(8) *Vallis Liliorum.*
(9, 10) *Sermones ad Novicios.*

The first five form sample A and the second five sample B of Table F. Sample A in this instance has more very short sentences, of ten words or less, than sample B, but the two are otherwise very much alike, and also resemble the distributions of Table E for the *Imitatio*. More exact comparison by the means, quartiles, etc., may be postponed till we make the summary comparison with the works of Gerson also. It is a small matter, but it may be mentioned that the "texts" of sermons were omitted

The edition of the works of Gerson that I used (ref. 9) is in four parts folio, and a selection for a sample had to be made from this rather appalling mass, a duty which could have been better performed by someone less ignorant of his work than myself. I tried to scatter the ten subsamples of about 120 sentences well over the four parts, to avoid matter that seemed hardly continuous prose or very exceptional in style and to choose matter that, in title at least, might not be too remote from something that Thomas might have treated. To reject something as "exceptional in style" may seem a dangerous proceeding, but I have in mind actually only one particular rejection, that of *De Modo Vivendi*

Omnium Fidelium. I put this down at first from its title but threw it out after examination. It consists of a series of brief rules, stated in curt sentences, after this style:

> Regula virginum. Non sint loquaces, sed simplices corde et habitu. Ad virginitatem matris Christi cogitent et eam diligant. Choreas vitent. Inter iuuenes non sedeant, nec se ab eis palpari permittant. Non ament aliquem illicito amore. Adulatores neque adulatrices recipiant nec audiant. Orationes libenter dicant. Sordida verba et inhonesta fugiant.

I hope it will be agreed that this is not normal prose—there is no continuity of thought nor development of ideas—but an exceptional *tour de force*, and was legitimately rejected. My subsamples were taken from the following:

(1) *Sermo factus in die circumcisionis Domini coram Papa apud Tarasconem.*

(2) *Tractatus contra sectum flagellantium se.* (A bad choice, as it is impossible to imagine Thomas à Kempis choosing such a subject.) As this proved too brief to give 120 sentences, sufficient was added from *Tractatus de probatione spirituum.*

(3) *Tractatus de parvulis trahendis ad Christum.*

(4) *Sermo de vita clericorum.*

(5, 6, 7) *De consolatione theologiae.* This is modelled on Boethius, *De consolatione philosophiae.* The three subsamples were taken from the beginning, middle and end. Verse was of course omitted.

(8) *De meditatione cordis*: the whole. As this gave only 109 sentences, on my reckoning, the deficiency was made up on the next two.

(9) *Sermo de circumcisione.*

(10) *Tractatus de consolatione in mortem amicorum.*

The first five form sample A of Table G, the second five sample B. It will be seen that the two are almost remarkably consistent with one another. I should add that I found the sentence difficulty distinctly troublesome at times with this edition of Gerson: full stops seem used too frequently and other punctuation marks inadequately. This impression was confirmed by the comparison mentioned in section I.

Finally, I decided to try an experiment with a different technique, pitching on columns by a random process and taking a sample of the same number of sentences from each. The parts or volumes I was using are numbered by columns, and the numbers of columns in these several volumes are as follows:

I. 934	III. 1190
II. 878	IV. 982

a total of nearly 4000 columns. Eliminating for simplicity the last 191 columns of Part III, any column can be specified by a number under 5000, the first digit giving the number of the Part, the last three digits the column; thus 2625 gives col. 625 of Part II, 4063 col. 63 of Part IV. Sequences of four consecutive numbers beginning with a 1, 2, 3, or 4 were then extracted from Tippett's *Random Numbers* and these taken as determining columns for samples. Numbers beyond the limits given above for Parts I, II and IV were simply dropped. But

numbers might also be rejected for other reasons: (1) the column might be verse; (2) it might contain matter not by Gerson at all, or only doubtfully by him; (3) the matter might be deemed otherwise unsuitable, i.e. hardly ordinary prose (cf. the rejection on the first sampling). I found it in fact quite impossible altogether to avoid the element of personal judgement and doubt now if it was desirable to attempt it: the point is discussed at the end of section IV. Relatively little was, however, rejected under the last head and the ground covered was, I think, more varied than before. When the column was fixed, I started with the first sentence beginning therein and continued straight ahead until 20 sentences had been counted. Samples A and B of Table H are therefore founded on 30 such "random passages" each, and the total column on 60 "random passages". If the "total" columns of Tables G and H are compared, it will be seen that they are closely similar.

If now Tables E and F for the *Imitatio* and the admitted miscellaneous works of à Kempis are compared with the Tables G and H for Gerson, it will be seen that there are very considerable differences, especially in the numbers of long or moderately long sentences, e.g. of more than 50 words. In Tables E and F these number 15 and 22 respectively; in Tables G and H they total to 68 and 66. For facility of checking, frequency distributions were booked up in the subsamples of about 120 sentences, and it is natural to enquire how far such small subsamples show consistent differences: it is obvious that no high degree of consistence is to be expected. The following are the numbers of sentences of 51 words or more in the subsamples of à Kempis and Gerson respectively, ranked in order of magnitude:

à Kempis: 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 3, 4, 5, 6, 7.
Gerson. 1, 2, 2, 3, 4, 5, 5, 5, 6, 6, 6, 7, 8, 8, 9, 10, 11, 11, 12, 13.

The upper quartile for Thomas à Kempis is 2·5, and this is exceeded by 17 of the 20 subsamples for Gerson. Seven of the subsamples for Thomas have no sentences at all of such a length: there is no subsample from Gerson without at least one. In both the range of variation exceeds, as one would expect, the value that would be given by the theory of simple sampling. On that theory the variance should be approximately equal to the mean, but the means and variances are:

à Kempis: M, 1·85; σ^2, 4·33
Gerson: M, 6·70; σ^2, 11·61

Roughly, fluctuations of simple sampling account for about half the variance in each case.

The complete comparison by means, quartiles, etc. is given in Table II. Comparing first the constants for the miscellaneous works of Thomas à Kempis with those for the *Imitatio*, and looking at the columns for samples A and B in both cases, we see that the values of the means overlap, that for sample A of the

TABLE II

Constants for the distributions of sentence-length in samples from the Imitatio Christi, *from Miscellaneous admitted works of Thomas à Kempis, and from Gerson. (Tables E, F, G and H of Appendix. Q_1 = Lower Quartile, Q_3 = Upper Quartile, D_9 = Ninth Decile)*

Constant	*Imitatio Christi*			à Kempis: Misc.		
	A	B	Total	A	B	Total
Mean	17·0	15·4	16·2	16·6	19 3	17·9
Median	14·0	13·6	13·8	13·8	16·4	15·1
Q_1	10·6	9·5	10·1	9·7	11·9	10·6
Q_3	20·7	18·4	19 3	20·8	23·9	22·4
Q_3-Q_1	10·1	8·9	9 2	11·1	12·0	11·8
D_9	28·6	26·0	27·7	29·3	32·5	31·0
	Gerson: Selected			Gerson. Random		
	A	B	Total	A	B	Total
Mean	23·5	23·4	23·4	23·5	22·0	22·7
Median	19·4	19·9	19·6	19·3	18·4	18·9
Q_1	12·5	12·6	12·5	12·0	11·4	11·7
Q_3	32·0	30·4	31·3	30·9	27·9	29·5
Q_3-Q_1	19·5	17 8	18·8	18·9	16·5	17·8
D_9	45·3	43·1	44·0	43·5	43·5	43·5

Miscellanea lying between the two values for the *Imitatio*. The values for the median and for the lower quartile overlap similarly. For the upper quartiles, the lower value for the Miscellanea, viz. 20·8, only just exceeds the upper value for the *Imitatio*, viz. 20·7; and there is a similar but slightly greater difference in the case of the interquartile range and the ninth decile. In no case are the differences at all large. The two tables for Gerson show a very similar degree of consilience.

But comparison of the constants for the *Imitatio* and the Miscellanea of Thomas à Kempis with those for Gerson's works shows quite a different state of affairs. For the lower quartile alone the differences are not large nor consistent, the lower quartile for sample B of the "random passages" from Gerson lying within the range of the lower quartiles for the Miscellanea of à Kempis and the *Imitatio*. All the remaining constants in the lower part of Table II are consistently larger than those in the upper part, and the differences are the more conspicuous the more the value of the constant is affected by long sentences:

it is largest (11–19 words) for the ninth decile, and next largest (4–14 words) for the upper quartile.

These results are completely consonant with the view that Thomas à Kempis was, and Jean Charlier de Gerson was not, the author of the *Imitatio.*

SECTION IV. GRAUNT'S *OBSERVATIONS UPON THE BILLS OF MORTALITY* AND THE ECONOMIC WRITINGS OF SIR WILLIAM PETTY

The problem of the authorship of the *Observations upon the Bills of Mortality* is, in all probability, of more interest to readers of this *Journal* than that of section III. At the same time it cannot be treated so completely as the problem of that section, for we have no other and admitted works by John Graunt with which to make comparison: we can only compare the one work which is generally believed to be by him with the admitted works by Sir William Petty.

The edition that I used both for the *Observations* and for Sir William Petty's writings was the convenient edition of Hull (ref. 15). Graunt gave me a certain amount of trouble in delimiting sentences, but the trouble was far more serious with Petty. I should like to quote, but the editors might reasonably object to my quoting several sentences each two or three hundred words or so in length: I must therefore merely refer readers to the original for illustrations. The longest sentence (as I reckoned it) in the *Observations* is the first part of §4, Chapter VII (ref. 15, vol. II, pp. 370–1). Here it seemed to me that the colon after "above-mentioned" on line 11 of p. 371 should be replaced by a full stop. This still leaves the sentence one of 213 words. On the other hand it appeared to me that the next following full stop between "*Annum*" and "And" on line 15 ought to be a comma, making the resulting sentence 70 words. This seemed a fairly clear case.

Take for comparison the longest sentence (again, as I reckoned it) in the samples from Petty, quite a characteristic loosely organized sequence of paragraphs in Chapter IV of the *Political Arithmetic* (ref. 15, vol. I, pp. 295–6). I allowed this sentence to begin with the words "To which purpose", the initial words in the last paragraph at the foot of p. 295, in spite of the relative adjective; but all the nine paragraphs beginning with "The value" on p. 296 had, it seemed to me, to be reckoned as part of the sentence, for the last alone possesses a verb. The result is that the sentence, on my reckoning, only stops at the words "Eighty thousand pounds" which close the paragraph towards the foot of p. 296. This is, I think, a lenient and doubtful reckoning. The first paragraph beginning "To which purpose" might well be taken as merely a relative clause properly belonging to the preceding paragraph, the sentence really beginning with the words "Now the *Wealth* of every Nation" in that paragraph, replacing the colon preceding "Now" by a full stop. This would add another 71 words to the 257 as I reckoned it in my work. Moreover, the paragraph following my

terminal limit on p. 296 leads off with "Which computation": this then might also be reckoned as a relative clause forming part of the same sentence, right down to the concluding words "Forty Five Millions", and adding yet another 105 words. On this computation then I ought to have reckoned the sentence as one of 433 words! This may sound almost incredible, but the sentence would really be no more than an expansion of a construction like this:

> Now, the wealth of a nation consisting chiefly in its share of the foreign trade of the world, we have to consider whether the English or the French have the greater *per capita* share of that trade; to which purpose I have estimated that the total value of the exports from Great Britain and Ireland, America, Africa, the East Indies, etc. amounts to some ten million pounds, a computation sufficiently justified by the Customs returns with an allowance for smuggling etc.

There is a special *type* of difficulty that occurs repeatedly, and may be illustrated by § 11, Chapter VI of the *Treatise of Taxes* (ref. 15, vol. I, p. 56). The paragraph starts "The Inconveniences of the way of Customs, are, viz.", and there then follow four numbered paragraphs with different grammatical relations to the introductory clause, like this, to abbreviate greatly:

(1) That duties are laid upon [raw materials etc.].

(2) The great number of officers requisite.

(3) The great facility of smuggling by bribery, etc.

(4) The customs and duties amount to so little that some other way of levy must be practised together with it.

No. 1 obviously forms part of the sentence with the introductory clause. Nos. 2 and 3 are not sentences as they stand, and ought to have been counted in also I think, but no. 4 is an independent sentence. Actually I find that in this case I do not seem to have obeyed my own rule that a word-sequence, to form a sentence, must be a grammatically complete expression of a thought, and nos. 2 and 3 were reckoned separately: this was, I believe done in some similar cases also. Indeed judging from the few instances where I have looked again at my classification some time after the original work was done, I seem to have been usually too merciful rather than too severe in placing the limits of the sentence. Difficulties were far more frequent and more troublesome than with any author I had tackled, and made the work both tedious and unsatisfactory, for far too much was thrown on my personal judgement. Hull says (ref. 15, pp. lxvii–lxviii):

> Unfortunately the use of rash calculations grew upon Petty, and as was to be expected, he gives widely varying estimates of the same things. It must be added that he is frequently inaccurate in his use of authorities and careless in his calculations and upon at least one occasion he is open to suspicion of sophisticating his figures.

This is sufficiently severe but I would add that, in my opinion, Petty's literary style, more especially in his argumentative writing, is loose and slovenly, indeed at times hardly grammatical. It is difficult to dissociate such slovenliness in

writing from slovenliness of thought. Only in purely descriptive matter does his style take on quite a different complexion.

They have a great Opinion of Holy-Wells, Rocks, and Caves, which have been the reputed Cells and Receptacles of men reputed Saints. They do not much fear Death, if it be upon a Tree, unto which, or the Gallows, they will go upon their Knees toward it, from the place they can first see it. They confess nothing at their Executions, though never so guilty. In brief, there is much Superstition among them, but formerly much more than is now; for as much as by the Conversation of Protestants, they become asham'd of their ridiculous Practices, which are not *de fide*. As for the Richer and better-educated sort of them, they are such Catholicks as are in other places. (*Political Anatomy of Ireland*, Chap. XII: ref. 15, vol. I, pp. 199–200.)

That is both pithy and picturesque.

So much for the difficulties; and now let us turn to the data. Graunt's *Observations* form but a slim volume, and his sentences tend to be long: omitting all prefatory matter and the appendix, and also one or two passages with tabular matter that it seemed impossible to deal with in any other way, I obtained no more than 335 sentences in all. The distribution is shown in Table J of the Appendix. To give some notion of the consistence of the style throughout, I have also broken up the total into three approximately equal subsamples. These are so small, and the run of the figures inevitably so irregular, that no very close consilience can be expected, but the degree of consistence does not seem to be at all unsatisfactory, and is particularly close as regards the numbers of longish sentences.

For facility of comparison, I thought it would be convenient to make the samples from Petty of the same size, and so intended: but, owing to a small revision made later in the Graunt table on looking through the work again, the totals for Petty are 334 against the 335 for Graunt. Sample A was taken mainly from the *Political Arithmetic*, as the work most closely associated with his name by statisticians. But this gave me only 300 sentences, and 34 were added from the *Treatise of Taxes* to make up the desired total. Sample B was taken wholly from the *Treatise of Taxes*. The distributions are given in Table K of the Appendix, and it will be seen that they are on the whole very concordant, with the exception that A shows a larger proportion of sentences of excessive length. If comparison be made with Table J it is obvious that these samples from Petty contain a very much larger proportion of long sentences than the *Observations*. There are only 17 sentences of 101 words or more in Table J, 54 and 45 sentences of 101 words or more in samples A and B of Table K. It may be added that this difference shows itself even in small subsamples. In the subsamples A, B and C of Table J there are 7, 6 and 4 such sentences. In corresponding subsamples of 111 or 112 sentences for samples A and B of Table K there are 24, 19, 11, 11, 15 and 19.

When I had got so far, I thought it would be of interest to supplement samples A and B for Petty's writings by a sample of "random passages" taken

in the same sort of way as for Gerson in section III. Hull's edition, though in two volumes, is paged continuously and runs only to 621 pages apart from appendices, index, etc.: omitting prefatory matter, the text of the first item (the *Treatise of Taxes*) does not start till p. 18. Pp. 314–438 are occupied by Graunt, with blank pages, title pages etc. I accordingly determined "random pages" by extracting from Tippett's *Random Numbers* triplets of digits beginning with 0, 1, 2, ..., 6, but not exceeding 621, and omitting numbers between the limits 000–018 and 314–438. A considerable number of the pages so given had to be struck out as either being blank pages, or containing prefatory matter, titles, contents, etc., or something obviously unsuitable such as tabular or semi-tabular matter. Very few were struck out as otherwise unsuitable, the only condition imposed being that the text should be fairly continuous ordinary prose, even though prose containing a good many figures: the limits were left as wide as possible. On each of 33 pages accepted I counted ten sentences, starting with the first complete sentence on the page and continuing till ten had been counted. On a supplementary 34th page I counted only four such sentences, so as to make up 334 sentences in all. We are dealing here with a much smaller range of numbers than in the Gerson experiment, and repetitions may occur: in fact, of the 55 numbers of three digits which were retained as lying within my limits and of which 22 were subsequently struck out as impossible or unsuitable, two occurred twice (one being amongst the subsequent rejections) and one three times. Two or three pairs might have been expected: the one occurrence of a triplet was unlikely.

The data given by this experiment are shown in column C of Table K of the Appendix. It will be seen that the first part of this distribution differs quite appreciably from the corresponding portions of columns A and B, there being a larger number of short sentences. But the "tail" of long sentences does not differ greatly, there being 40 sentences of 101 words or more in column C against 54 in column A and 45 in column B. The main source of the divergence is mentioned below, and the value of the sample discussed.

Table III gives the brief summary comparison in terms of means, quartiles etc. Taking first the medians and lower quartiles, all the three medians for Petty are higher than the median for the total of the *Observations*, which is the comparable figure based on the same number of sentences, but the median for sample C of Petty is lower than the median for sample A (based on only 111 sentences) of Graunt. A precisely similar statement is true for the lower quartiles. All the other constants, means, upper quartiles, interquartile ranges and ninth deciles are consistently higher for Petty than for Graunt, and the differences, especially for upper quartiles and ninth deciles, quite considerable. The distributions for the two authors seem to me completely differentiated: or, to put it otherwise, the results confirm other evidence that the actual *authorship* of the *Observations* is not the same as that of the economic writings of Sir William

TABLE III

Constants for the distributions of sentence-length in Graunt's Observations *and in samples from Petty's Works. (Tables J and K of Appendix)*

Constant	Graunt				Petty		
	A	B	C	Total	A	B	C
Mean	50·1	45·5	46·9	47·5	66·1	60·2	56·3
Median	45·2	38·0	37·4	40·1	56·9	51·3	44·0
Q_1	31·2	23·8	26·3	26·8	36·1	34·7	29 0
Q_3	63 3	55·5	65·5	62·3	83·2	79·0	73·7
Q_3-Q_1	32·1	31·7	39·2	35·5	47·1	44·3	44·7
D_9	85·2	85·0	85·2	85·2	126 0	109·3	110·1

Petty. Lord Lansdowne remarked, in replying to Prof. Greenwood (ref. 18, sentence quoted in ref. 19); "For literary style, neither the Observations nor Petty's writings are conspicuous, but I have yet to learn what differences can be detected between them in this respect." Sentence-length is surely one characteristic of *literary* style, and the difference seems clear. In the wider sense of *style*, the sense in which *le style c'est l'homme même*, the *Observations* seem to me to differ wholly from Petty's writings: they suggest a man of quite a different type of mind and quite a different character. The evidence from sentence-length is interesting, but adds very little.

To return in conclusion for a moment to the method of "random passages" in relation to this method of investigation, let me deal first with the reason for the divergence of sample C for Petty's writings from the two samples A and B. The latter were taken wholly from the *Political Arithmetic* and the *Treatise of Taxes*. Examining my 33 samples of ten sentences each for sample C, I found that eight (including the triplet and the pair) which were remarkable for the proportion of short sentences all came from the *Political Anatomy of Ireland*. The distribution for these 80 sentences alone is totally different from that of sample A or sample B, the constants being as follows: mean, 34·8; median, 31·2; Q_1, 24·7; Q_3, 42·2; Q_3-Q_1, 17·5; D_9, indeterminate within the blank range 59·5–62·5, say 61. Why this difference? I have already mentioned the reason and illustrated it by a quotation from this very tract. The matter is *purely descriptive*, descriptive (in the samples concerned) of the religion, diet, clothes, language and manners of the people of Ireland, and of the Government, militia and defence of the country; and when Petty has only to describe and not to argue he can apparently write like a Christian.* The *Observations* being, I think one may say, mainly argumentative, this sample of "random passages" is not properly comparable with

* Webster and the *O.E.D.* concur in classifying this expression as "Colloq. or Slang". But after all the early Christians, judging from both gospels and epistles, *did* write in short sentences.

it: it does not deal "with the same sort of material in the same sort of way" to quote the phrase from the beginning of section II. Ludicrously enough there really is no tract of Petty's in which he *does* deal with the same sort of material in the same sort of way as Graunt, so the condition is strictly impossible of fulfilment: we did our best in taking samples from two tracts that were both argumentative, and these two samples were very fairly consistent with each other.

But this result raises the whole question of method: was I right in attempting something like random sampling at all? The notion that samples ought to be random is so firmly engrained in one's mind that it seems almost sacrilegious to object to the application of the rule in a particular case. But after all the problem surely is *not* whether a tract passing under the name of Jones does or does not resemble, in this particular characteristic, a *random* sample from the writings of Brown, but samples from Brown's writings dealing, so far as possible "with the same sort of material in the same sort of way". The method of "selected samples" is, from this standpoint, entirely justified and perfectly correct. A critic may, of course, object to the particular choice of selected samples (the particular choice in this section and the last for example): but the *method* is right, and preferable to the method of "random passages" as I used it—that is to say with as little restriction as possible in regard to matter and treatment.

But there is this to be said. In the first place, used as I used it, the method does serve in some degree as a control and perhaps a warning. It brings out very well the apparent (comparative) homogeneity of Gerson's style in respect of sentence-length, and the heterogeneity of Petty's. In combination with selected samples it better exhibits all the facts. In the second place it might be used differently, just as much care being taken in deciding whether to accept or reject a passage given by the random numbers as in the case of the "selected samples", but thereby obtaining a wider range of selection.

Further, there is a danger in random sampling to which possibly I have not paid sufficient attention, the risk of bias in sampling arising from the varying lengths of sentences and the fact that the series of sentence-lengths, in order as they occur, is not a random one. To take a simple but extreme example, suppose our book consisted of equal numbers of pages containing respectively 30 sentences of 15 words each, and 15 sentences of 30 words each. Actually then the book would contain two sentences of 15 words to one of 30 words. But if we proceeded by the method used for obtaining "random passages" from Petty, taking only a sample of 10 sentences from each page determined by Tippett's numbers, we would tend to get a sample containing equal numbers of sentences of the two lengths: the number of long sentences would be overweighted. The difficulty would be surmounted if we made the sample, not a fixed number of sentences, but a fixed length of matter, say one page: or, provided the pages in the book were arranged fairly at random, by making the sample long enough to

cover a number of pages, like my subsamples of about 120 sentences. In fact of course no real case is as simple or extreme as this, and actually it will be remembered that the "random passages" sample from Petty (sample C) gave *fewer* long sentences and *more* short sentences than samples A and B, though this is no proof that it was not in some degree biased in the direction indicated. Some possible processes of sampling might easily lead to extreme bias of this type. Suppose, for example, we decided to make a random sample of single sentences, determining the page and the number of a word on the page by random numbers, and taking the sentence in which this word happened to fall. Then, it seems to me, the chance of a sentence being "caught" for the sample would be directly proportional to its length; for a sentence of 10 words would have ten chances of being caught and a sentence of 40 words forty chances. (The difficulty is closely analogous to that of determining size of family by asking casual people as to the number of their brothers and sisters.) The risk is much lessened, in my opinion, by taking longish samples and, of course, if we are mainly concerned with comparisons and not absolute figures, is less important, for the bias is unlikely to be very different in the two authors compared by the same method. The whole question of the best method to use for random sampling is, however, worth further discussion. So far as my own experience goes, however, I am inclined to prefer the method first used, the method of selected passages of considerable length.

REFERENCES

SECTION I

(1) KER, W. P. (1925). *Collected Essays*, ed. Charles Whibley, 2, 131. London: Macmillan.

(2) McKERROW, R. B. (1928). *An introduction to bibliography for literary students*, p. 250. Oxford. Clarendon Press.

SECTION II

(3) BACON, F. (1888). *The moral and historical works of Lord Bacon*, introduction and notes by Joseph Devey. London· George Bell and Sons.

(4) COLERIDGE, S. T. (1817). *Biographia literaria; or biographical sketches of my literary life and opinions.* London: Rest Fenner.

(5) LAMB, C. & LAMB, M. (1905). *The works of Charles and Mary Lamb*, ed. E. V. Lucas. London: Methuen.

(6) MACAULAY, T. (1888). *Critical and historical essays.* London: Longmans, Green, Reader and Dyer.

SECTION III

(7) GROOTE, G. (1937). *The following of Christ: the spiritual diary of Gerald Groote.* Translated into English from original Netherlandish texts as edited by James van Ginneken, S J., of the Catholic University of Nymegen, by Joseph Malaise, S.J. New York: America Press.

(8) À KEMPIS, T. H. (1904–22). *Opera Omnia*, voluminibus septem edidit Michael Josephus Pohl. Friburgi Brisigavorum: sumptibus Herder.

(9) GERSONII, IOANNIS. Doctoris et Cancellarii Parisiensis (1606). *Opera*; multo quam antehac auctiora et castigatiora; inque partes quatuor distributa. Parisiis.

(10) CRUISE, F. R. (1887). *Thomas à Kempis.* London: Kegan Paul, Trench. (Part IV deals with the authorship controversy.)

(11) Cruise, F. R. (1898). *Who was the author of the Imitation of Christ?* London: Catholic Truth Society. (A brief epitome.)

(12) Wheatley, L. A. (1891). *The story of the Imitatio Christi.* London: Elliot Stock.

(13) De Montmorency, J. E. C. (1906). *Thomas à Kempis: his age and book.* London: Methuen.

(14) De Backer, le R. P. Augustin (1864). *Essai bibliographique sur le livre De Imitatione Christi.* Liége: Grandmont-Donders. (Nos. 3057–3301 are items "relatifs à la contestation sur l'auteur".)

Section IV

(15) Hull, C. H. (1899). *The economic writings of Sir William Petty, together with the observations upon the bills of mortality more probably by Captain John Graunt.* Cambridge: University Press.

(16) Lansdowne, the Marquis of (1927). *The Petty Papers.* London: Constable.

(17) Greenwood, M. (1928). "Graunt and Petty." *J.R. Statist. Soc.* **91**, 79.

(18) Lansdowne, the Marquis of (1928). *The Petty-Southwell Correspondence.* London: Constable.

(19) Greenwood, M. (1933). "Graunt and Petty, a restatement." *J.R. Statist. Soc.* **96**, 76.

(20) Willcox, W. F. (1938). "The Founder of Statistics." *Rev. Inst. Int. Statist.* (5), **4**, 321.

APPENDIX OF TABLES

These tables are all in the same form, showing the numbers of sentences having the length (in words) stated in the left-hand column, in a sample or samples from the source stated in the heading and more fully in the preceding text. Thus, in a sample taken from the first portion of Bacon's *Essays*, column A shows that there was only one sentence (out of 462) of a length between 1 and 5 words, 8 with a length between 6 and 10 words, 24 with a length between 11 and 15 words, and so on. Blank lines have been omitted in the tails of the tables to save space.

TABLE A

Bacon's Essays (1597–1625)

A, first half to end of XXVI. B, second half to end of LI

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	1	2	3	121–125	3	4	7
6– 10	8	8	16	126–130	2	3	5
11– 15	24	25	49	131–135	2	1	3
16– 20	22	23	45	136–140	1	2	3
21– 25	46	53	99	141–145	3	2	5
26– 30	43	42	85	146–150	—	1	1
31– 35	57	55	112	151–155	1	2	3
36– 40	38	37	75	—	—	—	—
41– 45	24	38	62	166–170	—	1	1
46– 50	31	25	56	—	—	—	—
51– 55	23	28	51	186–190	1	—	1
56– 60	25	21	46	191–195	—	—	—
61– 65	19	17	36	196–200	1	—	1
66– 70	12	13	25	—	—	—	—
71– 75	19	8	27	211–215	1	—	1
76– 80	7	11	18	—	—	—	—
81– 85	12	11	23	226–230	—	1	1
86– 90	6	7	13	231–235	—	1	1
91– 95	6	9	15	—	—	—	—
96–100	2	11	13	311–315	—	1	1
101–105	7	3	10				
106–110	9	3	12				
111–115	4	1	5				
116–120	2	4	6	Total	462	474	936

TABLE B

Coleridge, Biographia Literaria (1817)

A, *vol.* I *to p.* 134. B, *vol.* II, *pp.* 1–66 *and* 104–*end* (*p.* 182)

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	9	2	11	101–105	4	6	10
6– 10	21	37	58	106–110	2	2	4
11– 15	46	44	90	111–115	1	1	2
16– 20	46	49	95	116–120	5	1	6
21– 25	58	73	131	121–125	2	3	5
26– 30	64	56	120	126–130	1	1	2
31– 35	55	57	112	131–135	1	1	2
36– 40	51	52	103	136–140	—	—	—
41– 45	49	52	101	141–145	—	2	2
46– 50	39	37	76	146–150	1	2	3
51– 55	24	29	53	151–155	—	1	1
56– 60	22	23	45	156–160	—	1	1
61– 65	21	18	39	161–165	1	—	1
66– 70	20	17	37	166–170	—	—	—
71– 75	20	9	29	171–175	—	1	1
76– 80	10	6	16	—	—	—	—
81– 85	6	9	15	196–200	1	—	1
86– 90	7	7	14				
91– 95	9	4	13				
96–100	5	3	8				
				Total	601	606	1207

TABLE C

Charles Lamb, Elia (1823) *and* Last Essays of Elia (1833)

A, *Elia*: from beginning to middle of Mrs Battle's Opinions on Whist. B, *Last Essays*: Detached Thoughts *on* Books to Barbara S-- *inclusive*

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	29	30	59	81– 85	7	6	13
6–10	115	100	215	86– 90	3	—	3
11–15	111	100	211	91– 95	5	2	7
16–20	61	85	146	96–100	2	1	3
21–25	62	56	118	101–105	3	1	4
26–30	36	46	82	106–110	1	—	1
31–35	36	46	82	111–115	1	1	2
36–40	21	29	50	116–120	1	—	1
41–45	16	19	35	121–125	1	1	2
46–50	19	16	35	126–130	1	2	3
51–55	13	18	31	131–135	1	—	1
56–60	5	6	11	136–140	2	1	3
61–65	15	11	26	—	—	—	—
66–70	2	5	7	171–175	—	1	1
71–75	7	8	15				
76–80	3	8	11				
				Total	579	599	1178

TABLE D

Macaulay

A, from first portion of essay on Lord Bacon (1837). B, from first portion of essay on The Earl of Chatham (1844)

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	26	20	46	71– 75	4	—	4
6–10	100	104	204	76– 80	4	4	8
11–15	126	126	252	81– 85	2	—	2
16–20	89	111	200	86– 90	2	—	2
21–25	82	104	186	91– 95	—	1	1
26–30	51	57	108	96–100	1	1	2
31–35	26	35	61	101–105	1	—	1
36–40	29	39	68	106–110	—	—	—
41–45	16	22	38	111–115	1	—	1
46–50	10	14	24	116–120	—	—	—
51–55	12	8	20	121–125	1	—	1
56–60	9	3	12				
61–65	7	1	8				
66–70	2	—	2	Total	601	650	1251

TABLE E

Imitatio Christi

A, from Lib. I, II and IV. B, from Lib. III

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	8	31	39	51– 55	6	1	7
6–10	142	160	302	56– 60	1	1	2
11–15	201	175	376	61– 65	1	1	2
16–20	108	129	237	66– 70	1	1	2
21–25	72	47	119	71– 75	—	—	—
26–30	33	19	52	76– 80	—	1	1
31–35	23	19	42	—	—	—	—
36–40	11	9	20	106–110	1	—	1
41–45	3	5	8				
46–50	6	5	11	Total	617	604	1221

TABLE F

Miscellaneous admitted works of Thomas à Kempis

For details as to the sources of samples A and B see text

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	33	14	47	51–55	3	5	8
6–10	153	98	251.	56–60	1	2	3
11–15	165	168	333	61–65	2	—	2
16–20	100	117	217	66–70	—	2	2
21–25	65	72	137	71–75	1	1	2
26–30	40	57	97	76–80	—	1	1
31–35	22	35	57	81–85	1	—	1
36–40	6	14	20	86–90	—	1	1
41–45	10	9	19	91–95	1	1	2
46–50	5	7	12				
				Total	608	604	1212

TABLE G

Gerson, Opera. *Selected samples*

For details see text

No. of words	Sentences			No. of words	Sentences		
	A	B	Total		A	B	Total
1– 5	30	29	59	61– 65	7	4	11
6–10	85	81	166	66– 70	3	5	8
11–15	108	115	223	71– 75	2	—	2
16–20	101	90	191	76– 80	2	2	4
21–25	68	78	146	81– 85	—	1	1
26–30	46	66	112	86– 90	—	2	2
31–35	53	45	98	91– 95	1	2	3
36–40	28	32	60	—	—	—	—
41–45	28	25	53	111–115	1	—	1
46–50	22	19	41	—	—	—	—
51–55	14	8	22	131–135	—	1	1
56–60	7	6	13				
				Total	606	611	1217

TABLE H

Gerson, Opera. *Random passages*

For details see text

No of words	Sentences A	Sentences B	Sentences Total	No. of words	Sentences A	Sentences B	Sentences Total
1– 5	23	34	57	61– 65	6	5	11
6–10	99	97	196	66– 70	4	6	10
11–15	97	111	208	71– 75	3	2	5
16–20	105	98	203	76– 80	2	2	4
21–25	75	80	155	81– 85	1	1	2
26–30	48	53	101	86– 90	1	—	1
31–35	43	26	69	91– 95	1	1	2
36–40	32	33	65	96–100	1	—	1
41–45	25	16	41	—	—	—	—
46–50	19	20	39	121–125	1	—	1
51–55	6	9	15	126–130	1	—	1
56–60	7	6	13				
				Total	600	600	1200

TABLE J

Graunt's Observations upon the Bills of Mortality

A, B, C, first, second and third portions: the whole included apart from some omissions (see text)

No. of words	Sentences A	Sentences B	Sentences C	Sentences Total	No. of words	Sentences A	Sentences B	Sentences C	Sentences Total
1– 5	—	—	—	—	86– 90	2	4	2	8
6–10	3	2	7	12	91– 95	2	—	1	3
11–15	2	9	2	13	96–100	—	1	4	5
16–20	5	9	9	23	101–105	1	—	—	1
21–25	8	12	9	29	106–110	1	1	1	3
26–30	8	11	6	25	111–115	—	—	—	—
31–35	12	8	20	40	116–120	1	2	—	3
36–40	10	10	8	28	121–125	1	1	1	3
41–45	8	8	8	24	126–130	2	—	—	2
46–50	8	6	1	15	131–135	—	—	—	—
51–55	9	9	5	23	136–140	1	—	—	1
56–60	8	3	4	15	—	—	—	—	—
61–65	4	3	5	12	151–155	—	—	1	1
66–70	5	4	6	15	156–160	—	1	1	2
71–75	5	2	5	12	—	—	—	—	—
76–80	3	3	2	8	211–215	—	1	—	1
81–85	2	2	4	8					
					Total	111	112	112	335

TABLE K

Petty

A, *Political Arithmetic*, 300 sentences, with 34 added from the *Treatise of Taxes*. B, *Treatise of Taxes*. C, random passages (see text)

No. of words	Sentences			No. of words	Sentences		
	A	B	C		A	B	C
1– 5	1	1	—	131–135	5	1	2
6– 10	4	3	6	136–140	3	2	2
11– 15	3	8	13	141–145	2	4	3
16– 20	11	21	17	146–150	4	3	4
21– 25	16	17	26	151–155	2	1	2
26– 30	20	20	31	156–160	1	—	1
31– 35	26	16	30	161–165	1	3	—
36– 40	22	31	27	166–170	1	1	—
41– 45	18	28	24	171–175	1	1	—
46– 50	28	19	18	176–180	—	—	—
51– 55	12	18	11	181–185	1	—	—
56– 60	21	15	14	186–190	—	1	—
61– 65	23	14	16	191–195	—	—	—
66– 70	16	16	11	196–200	1	—	—
71– 75	10	13	10	201–205	—	—	—
76– 80	14	15	8	206–210	—	—	1
81– 85	10	7	12	211–215	2	1	2
86– 90	14	10	11	216–220	—	—	—
91– 95	6	9	5	221–225	1	—	1
96–100	5	8	4	226–230	—	—	—
101–105	3	4	2	231–235	1	—	1
106–110	5	10	5	236–240	—	—	—
111–115	3	2	1	241–245	1	—	—
116–120	4	8	5	—	—	—	—
121–125	5	3	5	256–260	1	—	—
126–130	6	—	3				
				Total	334	334	334

THE ESTIMATION OF THE LOCATION AND SCALE PARAMETERS OF A CONTINUOUS POPULATION OF ANY GIVEN FORM

By E. J. G. PITMAN, *University of Tasmania*

1. Introductory

In this paper we shall be concerned only with continuous chance variables which have "elementary" probability functions, i.e. if X is any chance variable considered, we shall assume that there exists a non-negative function $f(x)$, defined and continuous at almost every real value of x, such that the probability that X lies in any interval is equal to the integral of $f(x)$ over that interval. We shall call $f(x)$ simply the probability function of X.

The essential problem of estimation may be stated as follows. We have a sample consisting of n independently observed values of X,

$$x_1, x_2, \ldots, x_n.$$

The probability function of X, $f(x, \theta_1, \theta_2, \ldots)$, is of known form but involves certain parameters $\theta_1, \theta_2, \ldots$ whose values are not known, and we wish to estimate these values from an examination of the sample.

The sample may be specified by a point (the sample point) whose Cartesian co-ordinates are $(x_1, \ldots, x_n)$ in an n-dimensional space W (the sample space). For the co-ordinates of a variable point in this space we shall use

$$\xi_1, \xi_2, \ldots, \xi_n.$$

We shall write
$$F = \prod_{r=1}^{n} \{f(\xi_r, \theta_1, \theta_2, \ldots)\},$$

and call F the probability of the sample $\xi_1, \ldots, \xi_n$. Throughout the paper we shall denote by H a function of the ξ such that

$$E(H) = \int_W FH\, d\xi_1 \ldots d\xi_n$$

exists, E denoting expectation, or mean value. The points of W where F is not zero form a region which we shall denote by W_+; it will in general depend on the particular values of the θ. A line which contains internal points of W_+ will be said to belong to W_+.

This paper develops a general method of solving problems of estimation in which the unknown parameters are "location" or "scale" parameters. We suppose that the probability function of X is

$$\frac{1}{c} f\left\{\frac{x-a}{c}\right\},$$

and that the function $f(x)$ is known but that one or both of the parameters a, c, which determine respectively the location and the scale of the distribution of X, is unknown. This general problem has been considered by Fisher (1934, p. 303),

and the method of this paper is very closely related to Fisher's; but there is a difference in the approach to the problem, and perhaps also in the final point of view. Also, a number of questions not discussed by Fisher are dealt with here in detail. The approach to the problem is essentially on the lines of Neyman & Pearson (1936) and Neyman (1937), and I have purposely adopted a good deal of the notation and terminology of these writers.*

A probability function $f(x)$ which is such that $xf(x)$ remains bounded as x tends to ∞ or to $-\infty$ or to 0, will be said to possess the property κ_1. It is obvious that if this is the case, and if $0 \leqq m \leqq n-1$,

$$\int_{-\infty}^{\infty} t^m f(x_1-t)\dots f(x_n-t)\,dt$$

is convergent for all sets of values of the x when $f(x)$ is bounded, and for almost all sets of values when $f(x)$ is unbounded. In the latter case the values of $x_1, \dots, x_n$ could be so chosen that several of the functions $f(x_1-t), f(x_2-t), \dots$ would become infinite for some finite value of t, and this might prevent the convergence. By the substitution $v = 1/t$ we can show that, if $f(x)$ has the property κ_1,

$$\int_0^{\infty} \frac{1}{v^{n+1}} f\left\{\frac{x_1}{v}\right\}\dots f\left\{\frac{x_n}{v}\right\} dv$$

is convergent for all sets of values of the x when $f(x)$ is bounded, and for almost all sets of values when $f(x)$ is unbounded. If in addition $f(x)$ is bounded in the neighbourhood of 0, and $0 \leqq m \leqq n-1$,

$$\int_0^{\infty} \frac{v^m}{v^{n+1}} f\left\{\frac{x_1}{v}\right\}\dots f\left\{\frac{x_n}{v}\right\} dv$$

is convergent for all values of the x when $f(x)$ is bounded, and almost all values of the x when $f(x)$ is unbounded (but still bounded in the neighbourhood of 0). If $f(x)$ is a monotonic function of x when $|x|$ is sufficiently large, and is also either bounded in the neighbourhood of 0 or monotonic on each side of 0, it will possess the property κ_1, for, as is well known, from the convergence of

$$\int_{-\infty}^{0} f(x)\,dx \quad\text{and}\quad \int_0^{\infty} f(x)\,dx$$

it follows that $xf(x)$ tends to 0 as x tends to ∞ or $-\infty$ or 0. Thus all ordinary probability functions have this property.

* I do not agree with the statement (Neyman, 1938, p. 158) that the theory of confidence intervals and the theory of fiducial probability are two different things, and I hope that this paper may help to show that they are essentially the same and that their two points of view are both necessary for a full comprehension of the theory of estimation.

The relation between direct and inverse methods in statistics has been discussed by Jeffreys (1937). With the proper *a priori* probability distribution of a parameter, the results of the two methods are formally similar; but, of course, essentially different problems are being dealt with by the two "methods". However, the properties of "estimators", with which the present paper is largely concerned, are true whichever problem is being discussed.

[For some comment on what appears to be a real difference between Neyman's theory of confidence intervals and the approach of the present paper, see a Note in the Miscellanea section below. ED.]

If $x \log |x| f(x)$ remains bounded as x tends to ∞ or $-\infty$ or 0, we shall say that $f(x)$ possesses the property κ_2. By means of the substitution $t = -\log u$, we can show that

$$\int_{-\infty}^{\infty} t^m e^{-nt} f(x_1 e^{-t}) \dots f(x_n e^{-t})\, dt,$$

where $0 \leqq m \leqq n-1$, is convergent for all sets of values of the x when $f(x)$ is bounded, and for almost all sets of values when $f(x)$ is not bounded.

2. The estimation of a

Here we take the probability function of X as

$$f(x-a)$$

and a is to be estimated. In accordance with the notation of § 1 we write

$$F = f(\xi_1 - a) \dots f(\xi_n - a).$$

Make the change of co-ordinates,

$$\xi_1 = z_1,$$
$$\xi_r = z_1 + z_r \quad (r = 2, 3, \dots, n).$$

The Jacobian of the transformation is 1, so that over any part of W

$$\int FH\, d\xi_1 \dots d\xi_n = \int FH\, dz_1 \dots dz_n.$$

The locus, $z_2, z_3, \dots, z_n$, all constant,

is a straight line parallel to the line

$$\xi_1 = \xi_2 = \dots = \xi_n.$$

Any such line for which $\int_{-\infty}^{\infty} F\, dz_1 > 0$

will be denoted by L. The family of lines L will be the same for all values of a. A point $(x_1, \dots, x_n)$ which is on some L will be called an observable point.* We shall write

$$\frac{\int_{-\infty}^{\infty} FH\, dz_1}{\int_{-\infty}^{\infty} F\, dz_1} = E_L(H),$$

and call $E_L(H)$ the mean value of H on L. Since

$$E(H) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} FH\, dz_1 \dots dz_n = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left\{ E_L(H) \int_{-\infty}^{\infty} F\, dz_1 \right\} dz_2 \dots dz_n,$$

it is evident that if $E_L(H) = h$ (constant) for every L,

$$E(H) = h \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} F\, dz_1 \dots dz_n = h;$$

while if $E_L(H) > h$, $E(H) > h$.

* If there is no interval throughout which $f(x)$ is zero, all points will be observable; if $f(x)$ vanishes outside a certain finite interval some points will not be observable. Each L is the locus of the sample point corresponding to a given "configuration" (Fisher, 1934, p. 301).

If I' is a set of intervals on L we write

$$\frac{\int_{I'} F\,dz_1}{\int_{-\infty}^{\infty} F\,dz_1} = P\{I' \mid L\}.$$

Suppose that I' is determined on every L; denote by w' the region formed by all the I' and by $P\{w'\}$ the probability that the sample point will fall in w'. If $P\{I' \mid L\}$ has the same value α on every L, then $P\{w'\} = \alpha$, while if $P\{I' \mid L\} > \beta$ (constant), $P\{w'\} > \beta$. This can be proved in the same way, or it can be deduced from the previous result by defining H to have the value 1 at a point of an I', and the value 0 at any other point.*

If $(x_1, \ldots, x_n)$ is any fixed point on L, the co-ordinates $(\xi_1, \ldots, \xi_n)$ of any point on L may be expressed in the form

$$\xi_r - a = x_r - t \quad (r = 1, 2, \ldots, n).$$

We then have
$$\int_{-\infty}^{\infty} F\,dz_1 = \int_{-\infty}^{\infty} f(x_1 - t) \ldots f(x_n - t)\,dt$$

and
$$E_L(H) \int_{-\infty}^{\infty} f(x_1 - t) \ldots f(x_n - t)\,dt = \int_{-\infty}^{\infty} Hf(x_1 - t) \ldots f(x_n - t)\,dt.$$

Let I denote a set of non-overlapping intervals in $(-\infty, \infty)$. The points of the line L corresponding to values of t lying in I will form a set of intervals I' on L. We shall call I "proper" if its end-points,

$$A_1, A_2, \ldots$$

are functions of $x_1, \ldots, x_n$, not involving a, such that I' is independent of the particular fixed point $(x_1, \ldots, x_n)$ on L. The necessary and sufficient condition for this is obviously

$$A_r(x_1 + \lambda, \ldots, x_n + \lambda) \equiv A_r(x_1, \ldots, x_n) + \lambda \quad (r = 1, 2, \ldots).$$

It should be noted that $\pm\infty$ are suitable end-points. Throughout the discussion it is to be understood that I is proper. I depends on $x_1, \ldots, x_n$ but not on a; we express this by

$$I = I(x_1, \ldots, x_n).$$

On the other hand, I', which is independent of the particular point $(x_1, \ldots, x_n)$, does depend on a. Change of a from a_1 to a_2 will increase all the ξ co-ordinates of each of the end-points of I' by $a_2 - a_1$, so that I' will simply slide along L through a distance $(a_2 - a_1)\sqrt{n}$ in the positive direction (ξ_1 increasing) of L.

Let
$$P\{I' \mid L\} = \frac{\int_{I'} F\,dz_1}{\int_{-\infty}^{\infty} F\,dz_1} = \frac{\int_I f(x_1 - t) \ldots f(x_n - t)\,dt}{\int_{-\infty}^{\infty} f(x_1 - t) \ldots f(x_n - t)\,dt} = \alpha. \quad \ldots\ldots(1)$$

* It is assumed that I' varies with L in such a way that H is almost everywhere continuous. This is ensured in the applications.

In some of the applications of the theory, α is given and we have to determine I, in others, I is given and α is a function of it. If α were any given number between 0 and 1, we might determine I by (1) together with the requirement that the sum of the lengths of the intervals of I is to be a minimum. In general, I will then be uniquely determined; if so, it will be proper.

The points of L corresponding to values of t lying in I, i.e. the points of I', will be called points of acceptance. Since for the point $(x_1, \ldots, x_n)$ itself, $t=a$, the necessary and sufficient condition for this point to be a point of acceptance is that a lies in $I(x_1, \ldots, x_n)$, which we shall write

$$a \epsilon I(x_1, \ldots, x_n).$$

If points of acceptance are determined on every line L, they will form a region of acceptance $w'(a)$.* The remainder of the sample space will be called the critical region $w(a)$.† If α has the same value on every L, the probability, $P\{w'(a)\}$, that the sample point will fall in the region of acceptance is α, while if for every L, $\alpha > \beta$ (constant), the probability is greater than β. It will still be independent of the particular value of a.

The effect of a change in the value of a, say from a_1 to a_2, will be simply to move the region of acceptance, without change of form, through a distance $(a_2 - a_1)\sqrt{n}$ in the positive direction of the lines L. From this it easily follows that if I is chosen so that the sum of its lengths is a minimum for the corresponding α, and therefore the sum of the lengths of I' is also a minimum, then when $a = a_1$, the probability that the sample point falls in $w'(a_1)$ is greater than the probability that it falls in $w'(a_2)$. Using the notation and terminology of Neyman & Pearson (1936, p. 8), we have, with this choice of I,

$$P\{E \epsilon w'(a_1) \mid a_1\} > P\{E \epsilon w'(a_2) \mid a_1\},$$

and therefore $$P\{E \epsilon w(a_1) \mid a_1\} < P\{E \epsilon w(a_2) \mid a_1\},$$

Since $$P\{E \epsilon w(a_2) \mid a_2\} = P\{E \epsilon w(a_1) \mid a_1\},$$

this gives $$P\{E \epsilon w(a_2) \mid a_2\} < P\{E \epsilon w(a_2) \mid a_1\},$$

and so the critical region $w(a)$ is "unbiased". If the shortest I is not always unique, we shall have to replace the sign $<$ in the last statement by $\leqq$.

The relation between α and $I(x_1, \ldots, x_n)$ is

$$\int_I f(x_1 - t) \ldots f(x_n - t)\, dt = \alpha \int_{-\infty}^{\infty} f(x_1 - t) \ldots f(x_n - t)\, dt. \qquad \ldots\ldots(2)$$

It is convenient in practice to replace‡ the symbol t in (2) by the symbol a, and we write (2) in the form

$$k \int_I f(x_1 - a) \ldots f(x_n - a)\, da = \alpha, \qquad \ldots\ldots(3)$$

where $$\frac{1}{k} = \int_{-\infty}^{\infty} f(x_1 - a) \ldots f(x_n - a)\, da.$$

* Cf. Neyman (1937, p. 351). † Cf. Neyman & Pearson (1936, p. 5).

‡ This replacement could not have been made earlier without confusion; at this stage t is a mere dummy.

The definition of an observable point $(x_1, \ldots, x_n)$ now takes the form

$$\int_{-\infty}^{\infty} f(x_1 - a) \ldots f(x_n - a)\, da > 0.$$

We have seen that the necessary and sufficient condition for the point $(x_1, \ldots, x_n)$ to be a point of acceptance is

$$a \in I(x_1, \ldots, x_n).$$

When α is constant, the probability that the sample point $(x_1, \ldots, x_n)$ is a point of acceptance is α. Hence

$$P\{a \in I(x_1, \ldots, x_n)\} = \alpha.$$

If $\alpha > \beta$ (constant), we shall have

$$P\{a \in I(x_1, \ldots, x_n)\} > \beta.$$

We may sum up our results in the following theorem. If $I(x_1, \ldots, x_n)$ is proper, and defined for every observable point $(x_1, \ldots, x_n)$, and if

$$k \int_I f(x_1 - a) \ldots f(x_n - a)\, da = \alpha \text{ (constant)},$$

where

$$k \int_{-\infty}^{\infty} f(x_1 - a) \ldots f(x_n - a)\, da = 1,$$

then

$$P\{a \in I(x_1, \ldots, x_n)\} = \alpha,$$

while if

$$k \int_I f(x_1 - a) \ldots f(x_n - a)\, da > \beta \text{ (constant)},$$

$$P\{a \in I(x_1, \ldots, x_n)\} > \beta.$$

We shall express all this by saying that the fiducial function* for the estimation of a is

$$kf(x_1 - a) \ldots f(x_n - a).$$

We shall denote this function by $g(a)$.

The statement $$a \in I(x_1, \ldots, x_n) \qquad \ldots\ldots(4)$$

is a variable statement which is a function of $x_1, \ldots, x_n$. When particular, actually observed values of $x_1, \ldots, x_n$ are inserted in it, we obtain a definite statement about the unknown parameter a that is either true or false, and we shall not know which it is; but we do know that the probability that the variable statement (4), when used in this way, will give a true particular statement about a is α (supposed constant). As R. A. Fisher expresses it, the fiducial probability of the variable statement (4) is α. If we decide upon α, say 0·95, and then define I accordingly, we shall have a rule for automatically making a definite statement about the unknown parameter a whenever a set of values of the chance variable X is observed. A statistician using this rule can expect to be right about 95 times out of 100.

* I at first called it the fiducial probability function, but finally decided to shorten the name by dropping the word "probability". As will be seen later, problems of estimation can be dealt with completely, and very simply, by means of the fiducial function.

Suppose that
$$f(x) = 2x, \quad 0 \leqq x \leqq 1,$$
$$= 0, \quad x < 0 \text{ or } x > 1.$$

Let $x_1, \ldots, x_n$ be a sample from the (triangular) population with probability function $f(x-a)$. The distribution extends only from a to $a+1$. We shall denote the smallest and the largest of the x by x_S and x_L respectively. Since $f(x)$ vanishes outside the range (0, 1), the fiducial function for the estimation of a is

$$2^n k(x_1-a) \ldots (x_n-a),$$

if $a \leqq x_S$ and $x_L \leqq a+1$, i.e. if $x_L - 1 \leqq a \leqq x_S$, and is zero for all other values of a. Thus

$$\int_{-\infty}^{\infty} f(x_1-a) \ldots f(x_n-a)\, da = \int_{x_L-1}^{x_S} f(x_1-a) \ldots f(x_n-a)\, da.$$

Since the fiducial function vanishes outside the interval (x_L-1, x_S) and is monotonic decreasing in this interval, the shortest I will consist of a single interval with its lower end at $x_L - 1$. Thus the shortest I will be the interval (x_L-1, h), where

$$\int_{x_L-1}^{h} (x_1-a) \ldots (x_n-a)\, da = \alpha \int_{x_L-1}^{x_S} (x_1-a) \ldots (x_n-a)\, da,$$

that is
$$G(h) - G(x_L-1) = \alpha\{G(x_S) - G(x_L-1)\}, \qquad \ldots\ldots(5)$$

where
$$G(a) = \int_0^a (x_1-a) \ldots (x_n-a)\, da,$$

a polynomial of the $(n+1)$th degree in a. Thus (5) is an equation of the $(n+1)$th degree to determine h. It will have a single root in the range (x_L-1, x_S). With this value of h, the statement

$$x_L - 1 \leqq a \leqq h$$

has fiducial probability α.

If $g(a)$, the fiducial function for the estimation of a, is for all values of the x a unimodal function of a, i.e. if it is a strictly monotonic function of a in $b_1 \leqq a \leqq b_2$ and zero outside (b_1, b_2), or if it is strictly increasing in $b_1 \leqq a \leqq b_2$, strictly decreasing in $b_2 \leqq a \leqq b_3$, and zero outside (b_1, b_3), the shortest I will always be unique and will consist of a single interval. Any point of acceptance on L will have a greater probability (or likelihood) than any point on L which is outside I', and, whatever the value of a, I will include the maximum likelihood estimate of a. This is the value of a which makes

$$f(x_1-a) \ldots f(x_n-a)$$

a maximum. We shall denote it by

$$A_L(x_1, \ldots, x_n).$$

A sufficient condition for $g(a)$ to be unimodal for all values of the x is that $f(x)$ satisfy either of the following conditions:

(i) $f(x)$ strictly monotonic over a certain range of x and zero outside that range;

(ii) $\log f(x)$ a concave function of x over a certain range of x and $f(x)$ zero outside that range.

This is easily proved by using the relation

$$\log g(a) = \log k + \Sigma \log f(x_r - a),$$

and remembering that the sum of any number of strictly monotonic functions of the same type (increasing or decreasing) is strictly monotonic, the sum of any number of concave functions is concave, and that a concave function is unimodal. The normal, the gamma, the beta (except when U-shaped), the triangular, and the trapezoidal (except rectangular) distributions all have probability functions which satisfy (i) or (ii).

We have so far been discussing estimation by interval.* This is what is required in statistical tests; but in practice it is often necessary to decide on some definite number as our estimate of a. Any such estimate will be the value of some function of the sample values

$$A(x_1, \ldots, x_n),$$

which does not involve the unknown parameter a. If we have no source of knowledge of the value of a except the observed sample, any principle of estimation which would assign the value a_0 to a when the observed values of X were

$$x_1, x_2, \ldots, x_n,$$

would assign the value $a_0 + \lambda$ to a when the observed values were

$$x_1 + \lambda, \quad x_2 + \lambda, \quad \ldots, \quad x_n + \lambda.$$

The function A must therefore satisfy the relation

$$A(x_1 + \lambda, \ldots, x_n + \lambda) \equiv A(x_1, \ldots, x_n) + \lambda.$$

Any function which satisfies this relation will be called an estimator of a. We note that the end-points of a proper I must all be estimators, including in this category $\pm\infty$, which formally have the estimator property, $\pm\infty + \lambda = \pm\infty$. For a particular population, an estimator A will be a chance variable with a definite distribution. It is easy to see that for a population of given form the distribution of the chance variable $A - a$ is independent of the particular value of the population parameter a. The practical requirement is an estimator A whose distribution is such that it is not likely to differ very much from the true value of a.

For points on the line L through $(x_1, \ldots, x_n)$,

$$A(\xi_1, \ldots, \xi_n) = A(x_1 + a - t, \ldots, x_n + a - t) = A(x_1, \ldots, x_n) + a - t.$$

Hence on any line L the difference between two estimators is constant,

$$A_1(\xi_1, \ldots, \xi_n) - A_2(\xi_1, \ldots, \xi_n) = A_1(x_1, \ldots, x_n) - A_2(x_1, \ldots, x_n).$$

The fiducial function

$$g(a) = kf(x_1 - a) \ldots f(x_n - a)$$

is defined, non-negative, and integrable in $-\infty < a < \infty$ when $f(x)$ is bounded.

* Cf. Neyman (1937, p. 346).

When $f(x)$ is unbounded, the statement is true for almost all sets of values of the x. Hence, apart from any questions of probability, it may be looked on as the elementary frequency function of a continuous distribution. This distribution we shall call the fiducial distribution of a determined by $x_1, \ldots, x_n$. If $f(x)$ has the property κ_1 of § 1,

$$\int_{-\infty}^{\infty} a^{n-1} g(a)\, da = k \int_{-\infty}^{\infty} a^{n-1} f(x_1 - a) \ldots f(x_n - a)\, da$$

exists for all, or almost all, values of the κ, and the moments of the fiducial distribution, up to the $(n-1)$th at least, exist. If $\phi(a)$ is any function of a, we shall write

$$\int_{-\infty}^{\infty} \phi(a)\, g(a)\, da = E_g\{\phi(a)\}.$$

The mean value of $(A-a)^m$ on L is

$$\begin{aligned} E_L\{(A-a)^m\} &= k \int_{-\infty}^{\infty} \{A(\xi_1, \ldots, \xi_n) - a\}^m f(x_1 - t) \ldots f(x_n - t)\, dt \\ &= k \int_{-\infty}^{\infty} \{A(x_1, \ldots, x_n) - t\}^m f(x_1 - t) \ldots f(x_n - t)\, dt \\ &= \int_{-\infty}^{\infty} (A-a)^m g(a)\, da = E_g\{(A-a)^m\}, \end{aligned}$$

where it is to be understood that in the last line A means $A(x_1, \ldots, x_n)$. Similarly

$$E_L\{|A-a|^m\} = E_g\{|A-a|^m\}.$$

The mean, median, or any such point of the fiducial distribution is a function of the x which has the estimator property. This is so because an increase of each of the numbers $x_1, \ldots, x_n$ by the same number λ simply shifts the fiducial distribution, without change of form, through a distance λ in the positive direction.

We may take the median (assumed unique*) of the fiducial distribution as our estimator of a. We shall denote it by

$$A_C = A_C(x_1, \ldots, x_n).$$

Since
$$\int_{-\infty}^{A_C} g(a)\, da = \tfrac{1}{2},$$

the probability that
$$-\infty < a \leqq A_C$$

is $\frac{1}{2}$. Thus the median value of A_C is a. If A is any other† estimator,

$$\begin{aligned} E_L\{|A-a| - |A_C - a|\} &= E_L\{|A-a|\} - E_L\{|A_C - a|\} \\ &= E_g\{|A-a|\} - E_g\{|A_C - a|\} \\ &= 0 \text{ if } A = A_C \text{ on } L \\ &> 0 \text{ if } A \neq A_C \text{ on } L, \end{aligned}$$

* This will be so for all values of the x if, and only if, the distribution of X has no gaps. When the median estimator A_C is not unique, the theorems will still hold provided A is not a median estimator.

† Estimators which are identical for almost all values of the x are regarded as not different.

since the mean absolute deviation of the fiducial distribution is a minimum about its median A_C. Hence
$$E\{|A-a|-|A_C-a|\}>0,$$
that is
$$E\{|A-a|\}>E\{|A_C-a|\}. \qquad \ldots\ldots(6)$$

Thus A_C is the estimator with the smallest mean absolute error. It has another important and interesting property which entitles it to be called the "closest"* estimator of a. It is likely to be nearer to the true value of a than any other estimator; more precisely, the probability that
$$|A_C-a| \leqq |A-a|$$
is greater than $\frac{1}{2}$. Define I as $(-\infty, \infty)$ if $A(x_1, \ldots, x_n)$ is equal to $A_C(x_1, \ldots, x_n)$, and as the interval extending from $\frac{1}{2}(A_C+A)$ to ∞ or $-\infty$ which includes A_C if A and A_C are not equal at $(x_1, \ldots, x_n)$. In either case
$$\int_I g(a)\,da > \tfrac{1}{2}.$$
Hence
$$P\{a \in I\} > \tfrac{1}{2};$$
but
$$a \in I$$
implies
$$|A_C-a| \leqq |A-a|.$$

Another important estimator is A_M, defined by
$$A_M(x_1, \ldots, x_n) = E_g(a) = \int_{-\infty}^{\infty} a g(a)\,da.$$
Its mean value is a since
$$E_L(A_M-a) = E_g(A_M-a) = A_M - E_g(a) = 0,$$
and therefore
$$E(A_M-a) = 0.$$
By the method used to establish (6) we can show that it is the estimator with the smallest mean square error
$$E\{(A_M-a)^2\} < E\{(A-a)^2\}. \qquad \ldots\ldots(7)$$
The expression on the left-hand side of (7) is the variance of A_M; but the right-hand expression is not the variance of A unless $E(A)=a$. However, we can prove that not only is (7) true, but also the variance of A_M is less than the variance of A unless A_M-A is constant. If $E(A)=a+h$, replace A in (7) by $A-h$, and the result follows. If the chance variable X has a finite standard deviation, σ, the variance of the sample mean,
$$\bar{x} = (\Sigma x_r)/n,$$
is σ^2/n. Since $\bar{x}$ has the estimator property, this implies
$$E\{(A_M-a)^2\} < \sigma^2/n,$$
unless $A_M-\bar{x}$ is constant.

* Cf. Pitman (1937). At the time of writing that paper I had not thought of using the word "estimator" to make a clear distinction between the function of the sample values and its value in a particular observation, which is what we take as our "estimate" of a.

If we define $A_{(r)}$ by

$$\int_{-\infty}^{\infty} | A_{(r)} - a |^r g(a)\, da \quad \text{a minimum},$$

$A_{(r)}$ will be the estimator with the smallest mean rth power absolute error

$$E\{| A_{(r)} - a |^r\} < E\{| A - a |^r\}.$$

The maximum likelihood estimator A_L, mentioned above, is defined by $g(A_L)$ a maximum, and its value is the abscissa of the mode of the fiducial distribution. The mode of its distribution is a, and it is always included in the shortest I. Except in simple cases like the normal and exponential populations, its approximate numerical value will usually be easier to determine than that of any of the other estimators discussed above. Apart from these it seems to have no special advantages.

For the normal population

$$f(x-a) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right]$$

and

$$g(a) = \frac{\sqrt{n}}{\sigma\sqrt{(2\pi)}} \exp\left[-\frac{n(a-\bar{x})^2}{2\sigma^2}\right], \quad \bar{x} = \Sigma x_r/n,$$

where σ is supposed to be known. The fiducial distribution of a is normal with mean $\bar{x}$ and standard deviation $\sigma/\sqrt{n}$. In this case the estimators discussed above all coincide,

$$A_C = A_M = A_L = A_{(r)} = \bar{x}.$$

In this case $\bar{x}$ is the "best" estimator of a, the "best"* estimator being defined as follows. An estimator A_B is the best estimator of a if, for *all* positive values of h,

$$P\{| A_B - a | \leqq h\} \geqq P\{| A - a | \leqq h\},$$

and, for *some* positive values of h,

$$P\{| A_B - a | \leqq h\} > P\{| A - a | \leqq h\}.$$

If, for all values of the x, the fiducial distribution of a is symmetrical and also unimodal in the wider sense, i.e. if $g(a)$ is a non-decreasing function of a at values of a below the centre of symmetry and consequently non-increasing above the centre, A_C is the best estimator, and

$$A_C = A_M = A_{(r)}.$$

The last part of this statement is obvious. The first part follows from the fact that

$$\int_{A_C-h}^{A_C+h} g(a)\, da \geqq \int_{A-h}^{A+h} g(a)\, da$$

for all positive values of h, and, when $A(x_1, \ldots, x_n)$ is not equal to $A_C(x_1, \ldots, x_n)$,

$$\int_{A_C-h}^{A_C+h} g(a)\, da > \int_{A-h}^{A+h} g(a)\, da$$

* It has been objected that the use of "best" to denote a particular kind of estimator is somewhat provocative; but I submit that an estimator which possesses the property of the definition is undeniably the best.

for some positive values of h. The condition that the fiducial distribution be symmetrical and unimodal in the wider sense for all values of the x is obviously also necessary for the existence of a best estimator.

The fiducial distribution of a determined by a sample from the rectangular population which extends from $a-\frac{1}{2}$ to $a+\frac{1}{2}$ is a rectangular distribution extending from $x_L-\frac{1}{2}$ to $x_S+\frac{1}{2}$, where x_S and x_L denote respectively the smallest and the largest member of the sample. $\frac{1}{2}(x_S+x_L)$ is the best estimator.

For the exponential population,

$$\begin{aligned} f(x-a) &= e^{a-x} \qquad & x \geqq a, \\ &= 0 & x<a, \end{aligned}$$

we have

$$\begin{aligned} g(a) &= ne^{n(a-x_S)} \qquad & a \leqq x_S, \\ &= 0 & a > x_S, \end{aligned}$$

where x_S is the smallest member of the sample. Here

$$A_L = x_S, \quad A_M = x_S - 1/n, \quad A_C = x_S - (\log 2)/n.$$

For the triangular population discussed earlier in this section, A_C is the value of h corresponding to $\alpha = \frac{1}{2}$,

$$A_L = x_L - 1, \quad A_M = \frac{G_1(x_S) - G_1(x_L-1)}{G(x_S) - G(x_L-1)},$$

where

$$G(a) = \int_0^a (x_1 - a) \dots (x_n - a)\, da, \quad G_1(a) = \int_0^a (x_1 - a) \dots (x_n - a)\, a\, da.$$

3. The estimation of c

Here the probability function of X is

$$c^{-1} f(x/c), \quad c > 0,$$

and

$$F = c^{-n} f(\xi_1/c) \dots f(\xi_n/c).$$

If X takes only positive values, we can reduce this to the previous case by considering the distribution of $\log X$ and putting

$$\log c = \gamma.$$

The probability function for the distribution of $\log X$ is then

$$e^{x-\gamma} f(e^{x-\gamma}),$$

and γ plays the part of a in the previous discussion. The results obtained apply to all cases; but we must establish them by a method which applies to chance variables taking both positive and negative values. As the analysis is similar to that in § 2, it will be given only in outline.

A function $C(x_1, \dots, x_n)$ whose value may be used as an estimate of c, i.e. a c estimator, must evidently satisfy

(i) $C(x_1, \dots, x_n) \geqq 0$,

(ii) $C(\lambda x_1, \dots, \lambda x_n) \equiv \lambda C(x_1, \dots, x_n), \quad \lambda \geqq 0;$

so that C must be a positive homogeneous function of the first degree in the x. Any function of this type will be called a c estimator. Its logarithm, G, which will be a γ estimator, will satisfy

$$G(\lambda x_1, \ldots, \lambda x_n) = G(x_1, \ldots, x_n) + \log \lambda, \quad \lambda \geqq 0,$$

and any function of this type will be called a γ estimator. Note that 0 and ∞ formally have the C property, while $\pm\infty$ have the G property.

A half line or ray with one end at the origin will be denoted by R if it belongs to W_+. Any point which lies on some R is called observable. We define the mean value of H on R by

$$E_R(H) = \frac{\int_0^\infty FHr^{n-1}\,dr}{\int_0^\infty Fr^{n-1}\,dr},$$

where $$r = \sqrt{(\Sigma \xi_r^2)},$$

the distance of the point $(\xi_1, \ldots, \xi_n)$ from the origin. If $E_R(H)$ has the same value h on every R, $E(H) = h$, and if $E_R(H) > h$ (constant), $E(H) > h$. This is easily proved by changing to spherical polar co-ordinates

$$x_1 = r \cos\theta_1,$$
$$x_2 = r \sin\theta_1 \cos\theta_2,$$
$$\cdots\cdots\cdots\cdots\cdots$$

and remembering that the Jacobian of the transformation is r^{n-1} multiplied by a function of the θ.

For a set of intervals I' determined on R, we define $P\{I' \mid R\}$ by

$$P\{I' \mid R\} \int_0^\infty Fr^{n-1}\,dr = \int_I Fr^{n-1}\,dr,$$

and we have, as before, $P\{w'\} = \alpha$ if $P\{I' \mid R\} = \alpha$ (constant) for every R, and $P\{w'\} > \beta$ if $P\{I' \mid R\} > \beta$ (constant), where w' is the region generated by the I', and $P\{w'\}$ is the probability that the sample point falls in w'.

If $(x_1, \ldots, x_n)$ is a fixed point on R, the co-ordinates $(\xi_1, \ldots, \xi_n)$ of any point on R may be expressed in the form

$$e^{-\gamma}\xi_r = e^{-t}x_r \quad (r = 1, 2, \ldots, n).$$

For points on R $$F = e^{-n\gamma} f(e^{-t}x_1) \ldots f(e^{-t}x_n),$$

also $$E_R(H) = \frac{\int_0^\infty He^{-nt} f(e^{-t}x_1) \ldots f(e^{-t}x_n)\,dt}{\int_0^\infty e^{-nt} f(e^{-t}x_1) \ldots f(e^{-t}x_n)\,dt}.$$

If I is a set of intervals in $(-\infty, \infty)$, the points of R corresponding to values of t lying in I will be called points of acceptance and will form a set of intervals I'. I will be proper if I' is independent of the particular point $(x_1, \ldots, x_n)$ on R, the

necessary and sufficient condition for which is that the end-points of I be γ estimators (including possibly $\pm\infty$). The relation between I and $\alpha = P\{I' \mid R\}$, is

$$\int_I e^{-nt} f(x_1 e^{-t}) \dots f(x_n e^{-t})\, dt = \alpha \int_{-\infty}^{\infty} e^{-nt} f(x_1 e^{-t}) \dots f(x_n e^{-t})\, dt. \quad \dots\dots(8)$$

The points of acceptance on all the rays R form a region of acceptance $w'(\gamma)$, and the remainder of the sample space is the critical region $w(\gamma)$. The regions of acceptance corresponding to different values of γ will be similar and similarly situated, with the origin as centre of similarity. It can be shown that the critical region obtained by using on every R the shortest I for the corresponding α is unbiased. An observable point is one for which the integral on the right-hand side of (8) is not zero.

Finally we obtain this theorem. If $I(x_1, \dots, x_n)$ is proper and defined at every observable point, and if

$$k \int_I e^{-n\gamma} f(x_1 e^{-\gamma}) \dots f(x_n e^{-\gamma})\, d\gamma = \alpha \text{ (constant)},$$

where
$$k \int_{-\infty}^{\infty} e^{-n\gamma} f(x_1 e^{-\gamma}) \dots f(x_n e^{-\gamma})\, d\gamma = 1,$$

then
$$P\{\gamma \in I(x_1, \dots, x_n)\} = \alpha,$$

while if
$$k \int_I e^{-n\gamma} f(x_1 e^{-\gamma}) \dots f(x_n e^{-\gamma})\, d\gamma > \beta \text{ (constant)},$$

then
$$P\{\gamma \in I(x_1, \dots, x_n)\} > \beta.$$

Again we express all this by saying that the fiducial function for the estimation of γ is

$$g_1(\gamma) = k e^{-n\gamma} f(x_1 e^{-\gamma}) \dots f(x_n e^{-\gamma}),$$

and the continuous distribution with elementary frequency function $g_1(\gamma)$ is called the fiducial distribution of γ determined by $x_1, \dots, x_n$.

If
$$\gamma \in I(x_1, \dots, x_n)$$

is equivalent to
$$c \in J(x_1, \dots, x_n),$$

the end-points of the set of intervals J will be c estimators (including possibly 0 to ∞), and

$$k \int_I e^{-n\gamma} f(x_1 e^{-\gamma}) \dots f(x_n e^{-\gamma})\, d\gamma = k \int_J c^{-n-1} f(x_1/c) \dots f(x_n/c)\, dc.$$

J will be said to be proper for the estimation of c. The shortest I is determined by

$$\int_I d\gamma$$

a minimum for the corresponding α; hence the corresponding J makes

$$\int_J \frac{dc}{c}$$

a minimum. The fiducial function for the estimation of c is

$$g_2(c) = k c^{-n-1} f(x_1/c) \dots f(x_n/c), \quad c \geqq 0,$$

and the last theorem can be stated with $J, c, g_2(c)$ in place of $I, \gamma, g_1(\gamma)$ respectively.

The expression for the mean value of H on R is

$$E_R(H) = k\int_{-\infty}^{\infty} He^{-nt} f(x_1 e^{-t}) \ldots f(x_n e^{-t})\, dt = \int_{-\infty}^{\infty} H g_1(t)\, dt.$$

For points on the ray R through $(x_1, \ldots, x_n)$,

$$\begin{aligned} G(\xi_1, \ldots, \xi_n) - \gamma &= G(\xi_1 e^{-\gamma}, \ldots, \xi_n e^{-\gamma}) = G(x_1 e^{-t}, \ldots, x_n e^{-t}) \\ &= G(x_1, \ldots, x_n) - t, \end{aligned}$$

where G is any γ estimator. Hence for any function ϕ

$$\begin{aligned} E_R\{\phi(G-\gamma)\} &= \int_{-\infty}^{\infty} \phi\{G(\xi_1, \ldots, \xi_n) - \gamma\}\, g_1(t)\, dt \\ &= \int_{-\infty}^{\infty} \phi\{G(x_1, \ldots, x_n) - t\}\, g_1(t)\, dt \\ &= \int_{-\infty}^{\infty} \phi(G-\gamma)\, g_1(\gamma)\, d\gamma \\ &= E_g\{\phi(G-\gamma)\}, \qquad \ldots\ldots(9) \end{aligned}$$

where it is to be understood that in the last two lines G means $G(x_1, \ldots, x_n)$. In particular,

$$E_R\{(G-\gamma)^m\} = E_g\{(G-\gamma)^m\}$$

and

$$E_R\{|\,G-\gamma\,|^m\} = E_g\{|\,G-\gamma\,|^m\}.$$

The factor k in the expression for $g_1(\gamma)$ is evidently a homogeneous function of degree n in the x. Writing $g_1(\gamma)$ in the form $g_1(\gamma, x_1, \ldots, x_n)$ to indicate its dependence on the x, we have

$$g_1(\gamma + \log\lambda,\ \lambda x_1,\ \ldots,\ \lambda x_n) \equiv g_1(\gamma,\ x_1,\ \ldots,\ x_n).$$

Hence multiplying each of the numbers $x_1, \ldots, x_n$ by the same number λ will simply shift the fiducial distribution of γ, without change of form, through a distance $\log\lambda$ in the positive direction; therefore the mean, median, etc. of the fiducial distribution of γ all have the G property.

G_C, the median of the fiducial distribution, is the closest estimator of γ, and the estimator with the smallest mean absolute error. The median value of its distribution is γ.

$$G_M = E_g(\gamma) = \int_{-\infty}^{\infty} \gamma g_1(\gamma)\, d\gamma = \int_0^{\infty} (\log c)\, g_2(c)\, dc$$

will be the γ estimator with the smallest mean square error,

$$E\{(G_M-\gamma)^2\} < E\{(G-\gamma)^2\}.$$

Its mean value is γ. G_L, the maximum likelihood estimator, is defined as the value of γ which makes $g_1(\gamma)$ a maximum, and we can define $G_{(r)}$ as the estimator with the smallest mean rth power absolute error.

The mean, median, etc. of the fiducial distribution of c are c estimators; but the relations of the c estimators to one another are not as simple as those of the γ estimators. The median, C_C, is the closest estimator of c. Its median value is c,

and its logarithm is G_C; but it is not in general the c estimator with the smallest mean absolute error. Again, the mean value of the c estimator with the smallest mean square error is not c. These complications arise from the fact that the relation corresponding to (9) is

$$E_R\{\phi(C/c)\} = E_g\{\phi(C/c)\},$$

which is obtained from (9) by replacing $\phi(G-\gamma)$ by $\phi(e^{G-\gamma})$, $= \phi(C/c)$. Hence

$$\frac{E_R\{(C-c)^m\}}{c^m} = E_g\left\{\left(\frac{C-c}{c}\right)^m\right\} = \int_0^\infty \frac{g_2(c)}{c^m}(C-c)^m\,dc.$$

For the estimator with the smallest mean square error, we must have $E_R\{(C-c)^2\}$ a minimum, and therefore

$$\int_0^\infty \frac{g_2(c)}{c^2}(C-c)^2\,dc$$

is a minimum; hence $$\int_0^\infty \frac{g_2(c)}{c^2}(C-c)\,dc = 0,$$

that is $$CE_g(1/c^2) - E_g(1/c) = 0.$$

Thus $C_{(2)}$, the c estimator with the smallest mean square error, is defined by

$$C_{(2)} = \frac{E_g(1/c)}{E_g(1/c^2)}.$$

Since $$E_R(C-c) = cE_g\left\{\frac{C-c}{c}\right\} = c\{CE_g(1/c)-1\},$$

$$E_R(C_{(2)}-c) = c\left\{\frac{\{E_g(1/c)\}^2 - E_g(1/c^2)}{E_g(1/c^2)}\right\} < 0,$$

and therefore $$E(C_{(2)}-c) < 0.$$

A sufficient condition for $$E(C^m) = c^m$$

is $$C^m = \frac{1}{E_g(1/c^m)}.$$

Before leaving the general theory we note that if $f(x)$ has the property κ_1 of § 1 and is bounded in the neighbourhood of 0, the first $n-1$ moments of the fiducial distribution of c are finite for all values of the x when $f(x)$ is bounded and for almost all values when $f(x)$ is unbounded, and that if it has the property κ_2, the first $n-1$ moments of the fiducial distribution of γ are finite for all, or almost all, values of the κ.

If X is normally distributed about 0 with standard deviation c, its probability function is

$$\frac{1}{c\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2/c^2}$$

and $$g_2(c) = \frac{k}{c^{n+1}}e^{-\frac{1}{2}S/c^2}, \quad c \geqq 0,$$

where $S = \Sigma x_r^2$. If h is any positive homogeneous function of degree 0 in the x, $C = \sqrt{(\frac{1}{2}S/h)}$ is a c estimator. Hence if we determine h so that

$$\int_C^\infty g_2(c)\,dc = \alpha \text{ (constant)}, \qquad \ldots\ldots(10)$$

we shall have $$P\{c \geqq C\} = \alpha,$$

that is $$P\{\tfrac{1}{2}S/c^2 \leqq h\} = \alpha. \qquad \ldots\ldots(11)$$

By the substitution $\frac{1}{2}S/c^2 = u$, (10) becomes

$$\frac{1}{\Gamma(\frac{1}{2}n)}\int_0^h e^{-u}u^{n/2-1}\,du = \alpha, \qquad \ldots\ldots(12)$$

so that h is constant. Looking at it the other way, we see that if h is any given positive number and α is determined by (12), then (11) is true. In other words, for a fixed normal population of mean 0 and standard deviation c, the chance variable $\frac{1}{2}S/c^2$ has a $\Gamma(\frac{1}{2}n)$ distribution, as is well known.

If $$\frac{1}{\Gamma(\frac{1}{2}n)}\int_{h_1}^{h_2} e^{-u}u^{n/2-1}\,du = \alpha, \qquad \ldots\ldots(13)$$

then $$P\{h_1 \leqq \tfrac{1}{2}S/c^2 \leqq h_2\} = \alpha,$$

that is $$P\{\tfrac{1}{2}S/h_2 \leqq c^2 \leqq \tfrac{1}{2}S/h_1\} = \alpha,$$

which gives $$P\{\tfrac{1}{2}\log(\tfrac{1}{2}S/h_2) \leqq \gamma \leqq \tfrac{1}{2}\log(\tfrac{1}{2}S/h_1)\} = \alpha.$$

Thus fiducial ranges for c^2 and γ can be determined for any given value of α. For the shortest range of γ, which gives an unbiased critical region, we must have

$$\tfrac{1}{2}\log(\tfrac{1}{2}S/h_1) - \tfrac{1}{2}\log(\tfrac{1}{2}S/h_2) \quad \text{a minimum,}$$

and therefore $$\log h_2 - \log h_1 \quad \text{a minimum.} \qquad \ldots\ldots(14)$$

From (13) $$e^{-h_2}h_2^{n/2-1}\,dh_2 - e^{-h_1}h_1^{n/2-1}\,dh_1 = 0,$$

and from (14) $$\frac{dh_2}{h_2} - \frac{dh_1}{h_1} = 0;$$

therefore $$e^{-h_2}h_2^{n/2} = e^{-h_1}h_1^{n/2}. \qquad \ldots\ldots(15)$$

The critical region corresponding to values of h_1, h_2 determined by (13) and (15) is unbiased.*

The estimators discussed above are all simply expressible in terms of S.

$$G_M = E_g(\gamma) = E_g\{\tfrac{1}{2}\log(\tfrac{1}{2}S/u)\} = \tfrac{1}{2}\{\log\tfrac{1}{2}S - E_g(\log u)\},$$

$$E_g(\log u) = \frac{1}{\Gamma(\frac{1}{2}n)}\int_0^\infty e^{-u}u^{n/2-1}\log u\,du = \frac{\Gamma'(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n)};$$

therefore $$G_M = \tfrac{1}{2}\left\{\log S - \log 2 - \frac{\Gamma'(\frac{1}{2}n)}{\Gamma(\frac{1}{2}n)}\right\}.$$

When n is large, this is approximately

$$\tfrac{1}{2}(\log S - \log 2 - \log\tfrac{1}{2}n) = \tfrac{1}{2}\log(S/n).$$

* Cf. Neyman & Pearson (1936, p. 19), where $\frac{1}{2}v_1$, $\frac{1}{2}v_2$ take the place of h_1, h_2.

Denote the median of the $\Gamma(m)$ distribution by $h(m)$; it is approximately equal to $m-\frac{1}{3}$. The fiducial median value of u, $=\frac{1}{2}S/c^2$, is $h(\frac{1}{2}n)$; hence

$$C_C = \sqrt{\{\tfrac{1}{2}S/h(\tfrac{1}{2}n)\}}$$

and $$G_C = \tfrac{1}{2}\log\{\tfrac{1}{2}S/h(\tfrac{1}{2}n)\}.$$

The closest estimator of c^2 is $$C_C^2 = \tfrac{1}{2}S/h(\tfrac{1}{2}n),$$

which is approximately $$S/(n-\tfrac{2}{3}).$$

The c^2 estimator usually employed is

$$S/(n-1);$$

its mean value is c^2.

The simplicity of this case arises from the fact that the fiducial distribution (of c or γ) depends on $x_1, \ldots, x_n$, only through the value of S. When S is fixed, the fiducial distribution is the same no matter what the individual values of the x may be. The important estimators and fiducial ranges are all functions of S only. S is what is called a sufficient statistic* for the estimation of c or γ. Other cases which are equally simple because of the existence of a sufficient statistic are the generalized gamma distribution,

$$c^{-1}f(x/c) = \frac{e^{-x/c}(x/c)^{m-1}}{c\Gamma(m)}, \quad x \geqq 0,$$
$$= 0, \quad x < 0,$$

and the rectangular distribution,

$$c^{-1}f(x/c) = c^{-1}, \quad 0 \leqq x \leqq c,$$
$$= 0, \quad x < 0 \text{ or } x > c.$$

The fiducial functions for the estimation of c are respectively

$$g_2(c) = \frac{ke^{-n\bar{x}/c}}{c^{mn+1}}, \quad c \geqq 0,$$

and $$g_2(c) = \frac{k}{c^{n+1}}, \quad c \geqq x_L.$$
$$= 0, \quad c < x_L.$$

The sufficient statistics are $\bar{x}$ and x_L.

While the existence of a sufficient statistic simplifies the mathematics and enables us to obtain explicit expressions for the important estimators and for the fiducial ranges, the methods of this and the preceding section are in no way dependent on this existence. When the sample values have been observed, the fiducial distribution is determinate, and it is theoretically possible to obtain the values of A_C, A_M, etc. or of G_C, G_M, C_C, etc., as the case may be, or the values of the end-points of the fiducial ranges I or J, to any required degree of accuracy. With small samples the labour would not be great. A practical process to deal with large samples would depend on a simple approximation to the fiducial distribution; but it is not proposed to discuss that aspect of the problem here.

* See Neyman & Pearson (1936, p. 117) and Pitman (1936).

4. The estimation of a and c

The probability function of X is assumed to be

$$\frac{1}{c}f\left\{\frac{x-a}{c}\right\},$$

the function $f(x)$ being known but the parameters a and c both unknown, and c positive. Thus

$$F=\frac{1}{c^n}f\left\{\frac{\xi_1-a}{c}\right\}\dots f\left\{\frac{\xi_n-a}{c}\right\}.$$

In practical problems of estimation, the chance variable X will be the measure of some physical quantity, and a and c will be the measures of quantities of the same kind as X. Hence any function of the observed values $x_1, \dots, x_n$ whose value may be used as an estimate of a, i.e. any a estimator, A, must be homogeneous of the first degree in the x.* Also, it must still satisfy the relation of § 2,

$$A(x_1+\lambda, \dots, x_n+\lambda)\equiv A(x_1, \dots, x_n)+\lambda.$$

Combining these two, we have

$$A\left\{\frac{x_1+\lambda}{\mu}, \dots, \frac{x_n+\lambda}{\mu}\right\}\equiv\frac{A(x_1, \dots, x_n)+\lambda}{\mu}.$$

Any function of this type will be called an a estimator and will be denoted by A. The probability function of the chance variable $X+k$, where k is a constant, will differ from the probability function of X, only in the value of a; hence any c estimator, C, in addition to being positive homogeneous of the first degree in the x, must also be invariant with respect to change of origin, and therefore

$$C\left\{\frac{x_1+\lambda}{\mu}, \dots, \frac{x_n+\lambda}{\mu}\right\}\equiv\frac{C(x_1, \dots, x_n)}{\mu}, \quad \mu\geqq 0.$$

Any such function will be called a c estimator.

The change of co-ordinates required is a combination of those used in §§ 2 and 3;

$$\begin{aligned}\xi_1&=\xi_1,\\ \xi_2&=\xi_1+r\cos\theta_1, \qquad (r\geqq 0)\\ \xi_3&=\xi_1+r\sin\theta_1\cos\theta_2,\\ &\dots\dots\dots\dots\end{aligned}$$

* This restriction was not made in § 2 for the following reason. From consideration of dimensions it is evident that the probability function of X in § 2 must really be of the form $\frac{1}{c}f\left\{\frac{x-a}{c}\right\}$, where c is the measure of some quantity of the same kind as the quantities whose measures are X and a; but since c was supposed to be known it was absorbed in the functional symbol f by writing the probability function in the form $f(x-a)$. All that can be said about the dimensions of an a estimator is that it must be homogeneous of the first degree in $c, x_1, \dots, x_n$, and this does not restrict its degree in the x only.

The relation of $\xi_2-\xi_1,\ \xi_3-\xi_1,\ \ldots,\ \xi_n-\xi_1$

to $r,\ \theta_1,\ \ldots,\ \theta_{n-2}$

is the relation of rectangular Cartesians to spherical polars in $n-1$ dimensions. The Jacobian of the transformation is

$$\frac{\partial(\xi_1, \xi_2, \ldots, \xi_n)}{\partial(\xi_1, r, \theta_1, \ldots, \theta_{n-2})} = r^{n-2}\Phi,$$

where Φ is a function of the θ only.

The locus $\theta_1, \theta_2, \ldots, \theta_{n-2}$, all constant,

is a (two-dimensional) half-plane with the line

$$\xi_1 = \xi_2 = \ldots = \xi_n \qquad \ldots\ldots(16)$$

as its edge. That the locus consists of a half-plane only can be seen as follows. Any (two-dimensional) plane through the line (16) consists of two half-planes which join along this line. These half-planes are distinguished from one another by the signs of

$$\xi_2-\xi_1,\ \xi_3-\xi_1,\ \ldots,\ \xi_n-\xi_1.$$

These signs do not change over one half-plane; but they all change as the point $(\xi_1, \ldots, \xi_n)$ moves from one half-plane to the other. When the θ are all fixed, the signs of

$$\xi_2-\xi_1,\ \xi_3-\xi_1,\ \ldots,\ \xi_n-\xi_1$$

are all fixed because r is positive, therefore the locus consists of a half-plane only. We denote any such half-plane by Q, and define the mean value of H on Q by

$$E_Q(H) = \frac{\int_Q FHr^{n-2}\,d\xi_1\,dr}{\int_Q Fr^{n-2}\,d\xi_1\,dr}. \qquad \ldots\ldots(17)$$

It is then easy to show that $E(H)=h$ if $E_Q(H)=h$ (constant) on every Q, and $E(H)>h$ if $E_Q(H)>h$ on every Q.

If D' is any region in Q, we define $P\{D' \mid Q\}$ by

$$P\{D' \mid Q\} = \frac{\int_{D'} Fr^{n-2}\,d\xi_1\,dr}{\int_Q Fr^{n-2}\,d\xi_1\,dr}. \qquad \ldots\ldots(18)$$

Obviously $P\{w'\}=\alpha$ if $P\{D' \mid Q\}=\alpha$ (constant) on every Q, and $P\{w'\}>\beta$ if $P\{D' \mid Q\}>\beta$ (constant) on every Q, where w' is the region in W formed by all the D', and $P\{w'\}$ is the probability that the sample point falls in w'. Since θ_1 is constant on Q, and

$$\xi_2 = \xi_1 + r\cos\theta_1,$$

we may write (17) and (18) in the more convenient forms*

$$E_Q(H) = \frac{\int_Q (\xi_2-\xi_1)^{n-2} FH\, d\xi_1 d\xi_2}{\int_Q (\xi_2-\xi_1)^{n-2} F\, d\xi_1 d\xi_2}$$

and
$$P\{D' \mid Q\} = \frac{\int_{D'} (\xi_2-\xi_1)^{n-2} F\, d\xi_1 d\xi_2}{\int_Q (\xi_2-\xi_1)^{n-2} F\, d\xi_1 d\xi_2}.$$

The co-ordinates $(\xi_1, \ldots, \xi_n)$ of any point on the half-plane Q† through the point $(x_1, \ldots, x_n)$ may be expressed in the form

$$\frac{\xi_r - a}{c} = \frac{x_r - u}{v} \quad (r = 1, 2, \ldots, n).$$

Since v is equal to $c(x_2-x_1)/(\xi_2-\xi_1)$, it will always be positive. Note that, at points on Q,

$$\frac{A(\xi_1, \ldots, \xi_n) - a}{c} = A\left\{\frac{\xi_1 - a}{c}, \ldots, \frac{\xi_n - a}{c}\right\} = A\left\{\frac{x_1 - u}{v}, \ldots, \frac{x_n - u}{v}\right\}$$
$$= \frac{A(x_1, \ldots, x_n) - u}{v}, \qquad \ldots\ldots(19)$$

and similarly
$$\frac{C(\xi_1, \ldots, \xi_n)}{c} = \frac{C(x_1, \ldots, x_n)}{v}. \qquad \ldots\ldots(20)$$

Since
$$\frac{\partial(\xi_1, \xi_2)}{\partial(u, v)} = \frac{c^2(x_2-x_1)}{v^3},$$

$$E_Q(H) = \frac{\int_0^\infty \int_{-\infty}^\infty Hf\left\{\frac{x_1-u}{v}\right\} \ldots f\left\{\frac{x_n-u}{v}\right\} \frac{1}{v^{n+1}}\, du\, dv}{\int_0^\infty \int_{-\infty}^\infty f\left\{\frac{x_1-u}{v}\right\} \ldots f\left\{\frac{x_n-u}{v}\right\} \frac{1}{v^{n+1}}\, du\, dv}.$$

Write
$$g(u, v) = kf\left\{\frac{x_1-u}{v}\right\} \ldots f\left\{\frac{x_n-u}{v}\right\} \frac{1}{v^{n+1}},$$

where k is defined by
$$\int_0^\infty \int_{-\infty}^\infty g(u, v)\, du\, dv = 1,$$

then
$$E_Q(H) = \int_0^\infty \int_{-\infty}^\infty Hg(u, v)\, du\, dv.$$

We may specify a pair of values of u, v by a point in a plane—the parameter plane ψ—whose Cartesian co-ordinates are (u, v). If D is a region in ψ, the points

* If Q happens to lie in the hyper-plane $\xi_1 = \xi_2$, we must replace ξ_2 by ξ_r, where $\xi_r - \xi_1$ is not zero at all points of Q.

† Q is the locus of the sample point corresponding to a given "configuration" (Fisher, 1934, p. 304).

of Q corresponding to points (u, v) lying in D will be called points of acceptance, and will form a domain D', it being understood that D is proper, i.e. that D' is independent of the particular point $(x_1, \ldots, x_n)$ on Q. This will be so if the boundary curves of D have equations of the form

$$\phi\left\{\frac{x_1-u}{v}, \ldots, \frac{x_n-u}{v}\right\} = 0.$$

In particular, the straight lines

$$u = A(x_1, \ldots, x_n)$$

and

$$v = C(x_1, \ldots, x_n)$$

are suitable boundary curves, as may be seen by writing their equations in the forms

$$A\left\{\frac{x_1-u}{v}, \ldots, \frac{x_n-u}{v}\right\} = 0$$

and

$$C\left\{\frac{x_1-u}{v}, \ldots, \frac{x_n-u}{v}\right\} = 1.$$

The necessary and sufficient condition for the point $(x_1, \ldots, x_n)$ to be a point of acceptance is

$$(a, c) \epsilon D(x_1, \ldots, x_n),$$

and the relation between D and $\alpha = P\{D' \mid Q\}$ is

$$\int_D g(u, v)\, du\, dv = \alpha.$$

Replacing the symbols u, v in this equation by a, c, we may state that the fiducial distribution of a and c is determined by the fiducial function

$$g(a, c) = kf\left\{\frac{x_1-a}{c}\right\} \ldots f\left\{\frac{x_n-a}{c}\right\} \frac{1}{c^{n+1}}.$$

This means simply that if $D(x_1, \ldots, x_n)$ is proper, and defined at every point $(x_1, \ldots, x_n)$, and if

$$\int_D g(a, c)\, da\, dc = \alpha \quad \text{(constant)},$$

then

$$P\{(a, c) \epsilon D(x_1, \ldots, x_n)\} = \alpha,$$

while if

$$\int_D g(a, c)\, da\, dc > \beta \quad \text{(constant)},$$

then

$$P\{(a, c) \epsilon D(x_1, \ldots, x_n)\} > \beta.$$

The mean value theorems are obtained by using (19) and (20).

$$E_Q\left\{\phi\left(\frac{A-a}{c}\right)\right\} = \int_0^\infty \int_{-\infty}^\infty \phi\left(\frac{A(\xi_1, \ldots, \xi_n)-a}{c}\right) g(u, v)\, du\, dv$$

$$= \int_0^\infty \int_{-\infty}^\infty \phi\left(\frac{A-u}{v}\right) g(u, v)\, du\, dv = \int_0^\infty \int_{-\infty}^\infty \phi\left(\frac{A-a}{c}\right) g(a, c)\, da\, dc,$$

which we denote by

$$E_g\left\{\phi\left(\frac{A-a}{c}\right)\right\}, \qquad \ldots\ldots(21)$$

it being understood that in the last two lines A means $A(x_1, \ldots, x_n)$. Similarly

$$E_Q\{\phi(C/c)\} = E_g\{\phi(C/c)\}.$$

Make the region D in the (a, c) plane consist of a strip or set of strips parallel to the c-axis and extending from $c=0$ to $c=\infty$, with boundary lines,

$$a = A(x_1, \ldots, x_n).$$

The intersection of D with the a-axis is a set of intervals I whose end-points are a estimators. The expression for $\alpha = P\{D' \mid Q\}$ is now

$$\alpha = \int_I \int_0^\infty g(a, c)\, dc\, da = \int_I g_1(a)\, da,$$

where $$g_1(a) = \int_0^\infty g(a, c)\, dc.$$

The statement $$(a, c) \epsilon D$$

becomes $$a \epsilon I.$$

Hence the fiducial function for the estimation of a is

$$g_1(a) = k\int_0^\infty f\left\{\frac{x_1 - a}{c}\right\} \ldots f\left\{\frac{x_n - a}{c}\right\} \frac{dc}{c^{n+1}}.$$

The mean, median, etc. of this fiducial distribution of a are a estimators; but, owing to the denominator c in (21), the relations of these estimators to one another are not in general as simple as the relations of the estimators in § 2. A_C, the median of the fiducial distribution of a, is obviously the closest estimator of a, and its median value is a; but it is not in general the estimator with the smallest mean absolute error.

For the estimator with smallest mean square error, we must have

$$E_Q\left\{\left(\frac{A - a}{c}\right)^2\right\}, \quad = \int_0^\infty \int_{-\infty}^\infty \frac{(A-a)^2}{c^2} g(a, c)\, da\, dc,$$

a minimum, which requires

$$\int_0^\infty \int_{-\infty}^\infty \frac{A - a}{c^2} g(a, c)\, da\, dc = 0,$$

that is $$AE_g(1/c^2) - E_g(a/c^2) = 0.$$

Thus the required estimator is $A_{(2)} = \dfrac{E_g(a/c^2)}{E_g(1/c^2)}$.

Its mean value is not necessarily a.

In the same way we can show that the fiducial function for the estimation of c is

$$g_2(c) = \int_{-\infty}^\infty g(a, c)\, da = k\int_{-\infty}^\infty f\left\{\frac{x_1 - a}{c}\right\} \ldots f\left\{\frac{x_n - a}{c}\right\} \frac{da}{c^{n+1}}.$$

Since the mean value theorem for c estimators,

$$E_Q\{\phi(C/c)\} = E_g\{\phi(C/c)\} = \int_0^\infty \left\{\int_{-\infty}^\infty \phi(C/c)\, g(a, c)\, da\right\} dc$$
$$= \int_0^\infty \phi(C/c)\, g_2(c)\, dc,$$

is the same as in § 3, the relations of the c estimators to one another will be the same

here as there. The properties of the γ ($=\log c$) estimators will be simpler than those of the c estimators. G_M, G_C, $G_{(r)}$, C_C have the same properties as in § 3, e.g.

$$G_M = E_g(\log c) = \int_0^\infty g_2(c) \log c\, dc$$

has mean value γ, and is the γ estimator with the smallest mean square error.

If a statement is to be made about both a and c with a given fiducial probability, we cannot simply combine the separate statements about a and c; we must use the fiducial function $g(a, c)$. This is what is required in statistical tests involving both a and c. When D is defined at every point, the points of acceptance form a region of acceptance $w'(a, c)$, and the remainder of the sample space is the critical region $w(a, c)$. Suppose now that α is fixed, and that $D(x_1, \ldots, x_n)$ is defined at every point by

$$\int_D g(u, v)\, du\, dv = \alpha,$$

$$\int_D \frac{1}{v} du\, dv = \int_D du\, d(\log v) \quad \text{a minimum;}$$

it can easily be shown that D so defined is proper. We take a random sample of n values of X and then make the statement that

$$(a, c) \epsilon D(x_1, \ldots, x_n). \qquad \ldots\ldots(22)$$

We know that the probability of making a true statement in this way is α, no matter what the actual values of a and c may be. Suppose further that the purpose of our observations is to test the hypothesis that a and c have certain specified values, $a = a_1$, $c = c_1$. If (a_1, c_1) does not lie in D as thus determined by the sample values, the statement (22) contradicts the hypothesis and we therefore reject the hypothesis. If (a_1, c_1) does lie in D, the hypothesis is not contradicted by (22) and we accept it. The probability of rejecting a hypothesis when it is actually true will be $1-\alpha$. In terms of the sample space,* the hypothesis $a = a_1$, $c = c_1$ is accepted if the sample point falls in the region of acceptance $w'(a_1, c_1)$, and rejected if the point falls in the critical region $w(a_1, c_1)$. If D is defined as above, it can be shown that the probability that the sample point falls in $w'(a_1, c_1)$ is a maximum when $a = a_1$, $c = c_1$, and therefore the hypothesis $a = a_1$, $c = c_1$ is more likely to be accepted when it is true than when it is false. The critical region determined in this way is unbiased. Further discussion of critical regions and of statistical tests associated with them must be reserved for another paper.

Applying the theory to a normal population with probability function,

$$\frac{1}{c\sqrt{(2\pi)}} e^{-\frac{1}{2}(x-a)^2/c^2},$$

we have
$$g(a, c) = \frac{k}{c^{n+1}} e^{-\frac{1}{2}(S+n(a-\bar{x})^2)/c^2},$$

* Using the ideas of Neyman & Pearson (1936) and Neyman (1937).

where $S = \Sigma(x_r - \bar{x})^2$, $\bar{x} = \Sigma x_r/n$. Hence

$$g_1(a) = \int_0^\infty g(a,c)\,dc = \frac{k'}{\{S/n + (a-\bar{x})^2\}^{\frac{1}{2}n}}.$$

The fiducial distribution of a is symmetrical, with its mode at $\bar{x}$. $A_C = \bar{x}$, and we can show that $A_{(r)} = \bar{x}$, but there is no need to do this, for we have already done it in § 2. The proof given there that $\bar{x}$ is the best estimator still holds good, for in § 2 we were comparing $\bar{x}$ with a wider class of estimators which included all the estimators of this section.

If h_1, h_2 are fixed numbers, $h_1 < h_2$,

$$P\{\bar{x} + h_1\sqrt{(S/n)} \leqq a \leqq \bar{x} + h_2\sqrt{(S/n)}\} = P\left\{h_1 \leqq \frac{a-\bar{x}}{\sqrt{(S/n)}} \leqq h_2\right\} = \frac{\int_{h_1}^{h_2} \frac{dz}{(1+z^2)^{\frac{1}{2}n}}}{\int_{-\infty}^{\infty} \frac{dz}{(1+z^2)^{\frac{1}{2}n}}},$$

which is "Student's" result. For a given value α of the last expression, the fiducial range of a will be shortest when $h_2 - h_1$ is least, i.e. when $h_1 = -h_2$. Thus if

$$\int_{-h}^{h} \frac{dz}{(1+z^2)^{\frac{1}{2}n}} = \alpha \int_{-\infty}^{\infty} \frac{dz}{(1+z^2)^{\frac{1}{2}n}},$$

then

$$P\{\bar{x} - h\sqrt{(S/n)} \leqq a \leqq \bar{x} + h\sqrt{(S/n)}\} = \alpha,$$

and this is the shortest fiducial range for given α.

For the estimation of c we have

$$g_2(c) = \int_{-\infty}^{\infty} g(a,c)\,da = \frac{k'}{c^n} e^{-\frac{1}{2}S/c^2}.$$

This is the same as for the normal population in § 3 except that S has a different meaning and n is replaced by $n-1$. Thus the estimation of c from a sample of n from a normal population of unknown mean is essentially the same as the estimation of c from a sample of $n-1$ from a population of known mean.

Suppose that

$$f(x) = \frac{1}{\Gamma(m)} e^{-x} x^{m-1}, \quad x \geqq 0, \quad m > 0,$$
$$= 0, \quad x < 0,$$

and that we have a sample of n from the generalized gamma population with probability function

$$\frac{1}{c} f\left\{\frac{x-a}{c}\right\}.$$

The expression for $g(a,c)$ is

$$g(a,c) = \frac{k}{c^{n+1}} e^{n(a-\bar{x})/c} \prod_{r=1}^{n} \left\{\left(\frac{x_r - a}{c}\right)^{m-1}\right\}, \quad a \leqq x_S,$$
$$= 0, \quad a > x_S,$$

where x_S is the smallest sample value. Integrating from 0 to ∞ with respect to c, we obtain

$$g_1(a) = \frac{k' \Pi\{(x_r - a)^{m-1}\}}{(\bar{x} - a)^{mn}}, \quad a \leqq x_S,$$
$$= 0, \quad a > x_S.$$

In the particular case of the exponential population, $m=1$, the probability function is

$$\frac{1}{c}e^{(a-x)/c}, \quad x \geqq a,$$

and

$$0, \quad x < a.$$

Hence
$$g(a,c) = \frac{k}{c^{n+1}}e^{n(a-\bar{x})/c}, \quad a \leqq x_S,$$
$$= 0, \quad a > x_S.$$

Further
$$g_1(a) = \frac{k'}{(\bar{x}-a)^n} = \frac{(n-1)(\bar{x}-x_S)^{n-1}}{(\bar{x}-a)^n}, \quad a \leqq x_S,$$
$$= 0, \quad a > x_S.$$

Since
$$\int_{-\infty}^{a} g_1(a)\,da = \frac{(\bar{x}-x_S)^{n-1}}{(\bar{x}-a)^{n-1}},$$
the median A_C is given by
$$\frac{(\bar{x}-x_S)^{n-1}}{(\bar{x}-A_C)^{n-1}} = \tfrac{1}{2}.$$

Thus the closest estimator of a is*

$$A_C = \bar{x} - 2^{1/(n-1)}(\bar{x}-x_S).$$

The estimator with the smallest mean square error is

$$A_{(2)} = \frac{E_g(a/c^2)}{E_g(1/c^2)},$$

which is easily shown to be

$$\bar{x} - (1+1/n)(\bar{x}-x_S).$$

The fiducial distribution of a is unimodal with its maximum at the upper end-point x_S. Hence the shortest fiducial range has its upper end at this point. Putting $z = (x_S - a)/(\bar{x}-x_S)$, we obtain

$$P\{x_S - h(\bar{x}-x_S) \leqq a \leqq x_S\} = (n-1)\int_0^h \frac{dz}{(1+z)^n} = 1 - \frac{1}{(1+h)^{n-1}}.$$

For the exponential population, the fiducial function for the estimation of c is

$$g_2(c) = \int_{-\infty}^{x_S} g(a,c)\,da = \frac{k'e^{-T/c}}{c^n},$$

where $T = n(\bar{x}-x_S)$, and the estimators and fiducial ranges are easily determined.

The location and scaling of the rectangular population with centre a and range c is simple and interesting; but there is no space here for further discussion of illustrative examples; it may just be remarked that $\frac{1}{2}(x_S+x_L)$ is the best estimator of a.

* Cf. Pitman (1937, p. 220).

5. THE ESTIMATION OF THE DIFFERENCE BETWEEN THE LOCATION PARAMETERS OF TWO POPULATIONS OF THE SAME FORM

Suppose that the probability functions of the chance variables X and Y are respectively

$$\frac{1}{c}f\left\{\frac{x-a}{c}\right\} \quad \text{and} \quad \frac{1}{c}f\left\{\frac{x-a-b}{c}\right\},$$

and that we wish to estimate b, the other parameters being also unknown. A pair of samples of values of X and Y,

$$x_1, x_2, \ldots, x_m$$
$$y_1, y_2, \ldots, y_n$$

may be specified by a point in $(m+n)$-dimensional space. For the Cartesian co-ordinates of a variable point in this space we shall use

$$\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n,$$

and we shall write $$F = \frac{1}{c^{m+n}}\prod_1^m f\left(\frac{\xi_r-a}{c}\right)\cdot\prod_1^n f\left(\frac{\eta_r-a-b}{c}\right).$$

A b estimator is any function which is homogeneous of the first degree in the x and y and which satisfies the relations

$$B(x_1+\lambda, \ldots, x_m+\lambda, y_1, \ldots, y_n) = B(x_1, \ldots, x_m, y_1, \ldots, y_n)-\lambda,$$
$$B(x_1, \ldots, x_m, y_1+\lambda, \ldots, y_n+\lambda) = B(x_1, \ldots, x_m, y_1, \ldots, y_n)+\lambda.$$

The transformation of §4 is applied separately to the ξ and η co-ordinates, with a slight modification for the latter:

$$\begin{aligned} \xi_1 &= \xi_1, & \eta_1 &= \eta_1 \\ \xi_2 &= \xi_1+r\cos\theta_1, & \eta_2 &= \eta_1+rs\cos\phi_1 \\ \xi_3 &= \xi_1+r\sin\theta_1\cos\theta_2, & \eta_3 &= \eta_1+rs\sin\phi_1\cos\phi_2, \\ &\cdots\cdots & &\cdots\cdots \end{aligned}$$

The Jacobian is $r^{m+n-3}s^{n-2}\Phi_1\Phi_2$, where Φ_1 is a function of the θ and Φ_2 a function of the ϕ only.

The locus $\quad s, \theta_1, \ldots, \theta_{m-2}, \phi_1, \ldots, \phi_{n-2}$, all constant,

is a three-dimensional half-space Q, bounded by the two-dimensional plane

$$\xi_1 = \xi_2 = \ldots = \xi_m, \quad \eta_1 = \eta_2 = \ldots = \eta_n.$$

The definitions of $E_Q(H)$ and $P\{D' \mid Q\}$ are*

$$E_Q(H) = \frac{\int_Q FHr^{m+n-3}\,d\xi_1 d\eta_1 dr}{\int_Q Fr^{m+n-3}\,d\xi_1 d\eta_1 dr} = \frac{\int_Q (\xi_2-\xi_1)^{m+n-3}\,FH\,d\xi_1 d\xi_2 d\eta_1}{\int_Q (\xi_2-\xi_1)^{m+n-3}\,F\,d\xi_1 d\xi_2 d\eta_1},$$

$$P\{D' \mid Q\} = \frac{\int_{D'} Fr^{m+n-3}\,d\xi_1 d\eta_1 dr}{\int_Q Fr^{m+n-3}\,d\xi_1 d\eta_1 dr} = \frac{\int_{D'} (\xi_2-\xi_1)^{m+n-3}\,F\,d\xi_1 d\xi_2 d\eta_1}{\int_Q (\xi_2-\xi_1)^{m+n-3}\,F\,d\xi_1 d\xi_2 d\eta_1},$$

* As before, if Q happens to lie in the hyper-plane $\xi_1=\xi_2$, we must replace ξ_2 by ξ_r, where $\xi_r-\xi_1$ is not zero at all points of Q.

where D' is a domain in Q, and it can easily be proved that $E_Q(H)$ and $P\{D' \mid Q\}$ have the same properties as before.

The co-ordinates $(\xi_1, \ldots, \xi_m, \eta_1, \ldots, \eta_n)$ of a point in the half-space through $(x_1, \ldots, x_m, y_1, \ldots, y_n)$ may be expressed in terms of three variables t, u, v, as follows:

$$\frac{\xi_r - a}{c} = \frac{x_r - t}{v} \quad (r = 1, 2, \ldots, m),$$

$$\frac{\eta_r - a - b}{c} = \frac{y_r - t - u}{v} \quad (r = 1, 2, \ldots, n).$$

Proceeding as before, we finally obtain as the fiducial function for the estimation of b,

$$g_1(b) = \int_{-\infty}^{\infty}\int_0^{\infty} g(a, b, c)\, dc\, da,$$

where
$$g(a,b,c) = \frac{k}{c^{m+n+1}} \prod_1^m f\left\{\frac{x_r - a}{c}\right\} \cdot \prod_1^n f\left\{\frac{y_r - a - b}{c}\right\},$$

and k is defined by
$$\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(a, b, c)\, da\, db\, dc = 1.$$

If X and Y are normal variables with the same standard deviation c, and with means a and $a+b$ respectively,

$$g(a,b,c) = \frac{k}{c^{m+n+1}} e^{-\frac{1}{2}T/c^2},$$

where
$$T = S_1 + S_2 + m(a - \bar{x})^2 + n(a + b - \bar{y})^2,$$

$$S_1 = \sum_1^m (x_r - \bar{x})^2, \quad S_2 = \sum_1^n (y_r - \bar{y})^2.$$

Hence
$$\int_0^{\infty} g(a, b, c)\, dc = k_1 T^{-\frac{1}{2}(m+n)}$$

$$= \frac{k_1}{\{S_1 + S_2 + m(a - \bar{x})^2 + n(a + b - \bar{y})^2\}^{\frac{1}{2}(m+n)}}.$$

Integrating this from $-\infty$ to ∞ with respect to a, we obtain

$$g_1(b) = \frac{k_2}{\{(S_1 + S_2)(1/m + 1/n) + (b - \bar{y} + \bar{x})^2\}^{\frac{1}{2}(m+n-1)}}.$$

Thus the fiducial distribution of b is of the same form as the fiducial distribution of a determined by a sample of $m+n-1$ from a single normal population of unknown mean and unknown standard deviation. We have finally

$$P\left\{\bar{y} - \bar{x} - h\sqrt{\left(\frac{(m+n)(S_1+S_2)}{mn}\right)} \leq b \leq \bar{y} - \bar{x} + h\sqrt{\left(\frac{(m+n)(S_1+S_2)}{mn}\right)}\right\}$$

$$= \frac{1}{B\{\frac{1}{2}, \frac{1}{2}(m+n-2)\}} \int_{-h}^{h} \frac{dz}{(1+z^2)^{\frac{1}{2}(m+n-1)}}.$$

Consider now two exponential populations with probability functions

$$c^{-1}e^{-(x-a)/c},\ x \geqq a \quad \text{and} \quad c^{-1}e^{-(y-a-b)/c},\ y \geqq a+b.$$

$$g(a,b,c) = \frac{k}{c^{m+n+1}}e^{-\{m(\bar{x}-a)+n(\bar{y}-a-b)\}/c}, \quad a \leqq x_S, \quad a+b \leqq y_S,$$

$$= 0 \text{ for all other sets of values of } a,\ b.$$

where x_S is the smallest x and y_S the smallest y.

Write
$$g(a,b) = \int_0^\infty g(a,b,c)\,dc;$$

then
$$g(a,b) = \frac{k_1}{\{m(\bar{x}-a)+n(\bar{y}-a-b)\}^{m+n}}, \quad a \leqq x_S, \quad a+b \leqq y_S,$$

$$= 0 \text{ for all other sets of values of } a,\ b.$$

The conditions $a \leqq x_S$, $a+b \leqq y_S$ are equivalent to

$$a \leqq x_S \text{ when } b \leqq y_S - x_S,$$

and
$$a \leqq y_S - b \text{ when } b \geqq y_S - x_S.$$

Put $B = y_S - x_S$, $C = m(\bar{x}-x_S)+n(\bar{y}-y_S)$; then when $b \leqq B$,

$$g_1(b) = \int_{-\infty}^{\infty} g(a,b)\,da = \int_{-\infty}^{x_s} \frac{k_1\,da}{\{m(\bar{x}-a)+n(\bar{y}-a-b)\}^{m+n}}$$

$$= \frac{k_1}{(m+n)(m+n-1)\{m(\bar{x}-x_S)+n(\bar{y}-x_S-b)\}^{m+n-1}}$$

$$= \frac{k_2}{\{C+n(B-b)\}^{m+n-1}}$$

Similarly, when $b \geqq B$, $\quad g_1(b) = \dfrac{k_2}{\{C+m(b-B)\}^{m+n-1}}.$

If h is positive

$$\int_{B-h}^{B} g_1(b)\,db = \frac{k_2}{n(m+n-2)}\left\{\frac{1}{C^{m+n-2}} - \frac{1}{(C+nh)^{m+n-2}}\right\}$$

and
$$\int_{B}^{B+h} g_1(b)\,db = \frac{k_2}{m(m+n-2)}\left\{\frac{1}{C^{m+n-2}} - \frac{1}{(C+mh)^{m+n-2}}\right\}.$$

Since
$$\int_{-\infty}^{\infty} g_1(b)\,db = 1,$$

this gives
$$k_2 = \frac{mn(m+n-2)\,C^{m+n-2}}{m+n}.$$

Hence, if h_1 and h_2 are positive,

$$\int_{B-h_1}^{B+h_2} g_1(b)\,db = 1 - \frac{m}{(m+n)(1+nh_1/C)^{m+n-2}} - \frac{n}{(m+n)(1+mh_2/C)^{m+n-2}}.$$

For a given value α of this integral, the range $(B-h_1,\ B+h_2)$ will be shortest when $g_1(B-h_1) = g_1(B+h_2)$, that is when

$$C+nh_1 = C+mh_2.$$

Putting $$nh_1 = mh_2 = pC,$$

we have $$\alpha = \int_{B-h_1}^{B+h_1} g_1(b)\,db = 1 - \frac{1}{(1+p)^{m+n-2}}.$$

Hence $$P\{B - pC/n \leqq b \leqq B + pC/m\} = \alpha,$$

where $$p = \frac{1}{(1-\alpha)^{1/(m+n-2)}} - 1.$$

Upon this result we can base a test for exponential populations analogous to Fisher's extension of "Student's" test for normal populations. A similar test for rectangular populations can be obtained in the same way.

6. Concluding remarks

More complicated problems of estimation of location and scale parameters, for example those which arise when we have samples from more than two populations, can be dealt with by the methods of this paper. Questions about statistical tests of hypotheses concerning such parameters can be treated in the same way. Here it has been impossible to do more than just glance at this side of the subject; but it is hoped to continue the discussion in a later paper.

Summary

The main problem considered is the location and scaling of the distribution of a continuous chance variable X. We suppose that the probability function of X is

$$\frac{1}{c} f\left\{\frac{x-a}{c}\right\}, \quad c > 0,$$

where the function $f(x)$ is known but one or both of the parameters a, c, which determine respectively the location and scale of the distribution, is unknown. We have a sample of n independently observed values of X, and from these we have to estimate the unknown parameter or parameters. Any function of the sample values whose value may be used as an estimate of an unknown parameter is called an estimator of that parameter. The paper shows how to determine an estimator with any required property, such as minimum mean absolute error, or minimum mean square error. In particular, the closest estimator is determined; this is an estimator whose median value is the true value of the parameter and which is likely to be closer to the true value than any other estimator. It is shown that in certain cases a best estimator exists.

Fiducial limits for the unknown parameter are determined, and what is called the fiducial distribution of the parameter is defined. It is shown that problems of estimation can be dealt with very simply, and completely, by means of fiducial distributions. For a population of any given form, the fiducial distribution of a, when both a and c are unknown, provides us with a test which corresponds to

"Student's" test for significance of the mean of a sample from a normal population.

The estimation of the difference between the location parameters of two populations of similar forms is discussed.

REFERENCES

FISHER, R. A. (1934). *Proc. Roy. Soc.* A, **144**, 285–307.
JEFFREYS, H. (1937). *Proc. Roy. Soc.* A, **160**, 325–48.
NEYMAN, J. & PEARSON, E. S. (1936). *Statist. Res. Mem.* **1**.
NEYMAN, J. (1937). *Philos. Trans.* A, **236**, 333–80.
—— (1938). *Lectures and Conferences on Mathematical Statistics.* Graduate School of the U.S. Dept. of Agriculture, Washington.
PITMAN, E. J. G. (1936). *Proc. Camb. Phil. Soc.* **32**, 567–79.
—— (1937). *Proc. Camb. Phil. Soc.* **33**, 212–22.

METHODS OF ESTIMATING THE POPULATION OF INSECTS IN A FIELD

By GEOFFREY BEALL

Dominion Entomological Laboratory, Chatham, Ontario, Canada

Purpose of study

Counts on the occurrence of an insect were secured to make clear a valid and efficient method of estimating the population of the insect in an area. The theory of sampling, as developed by Neyman (1934), was applied to this problem. It was desired to know: first, what form observations should take in sampling; secondly, how good are the results of stratification, or control of regional variability; and, thirdly, how accuracy varies when various fractions of the total area are sampled. The present study should supply a general method of sampling to be applied in experimental work or in surveys. Details in connexion with the method would, presumably, vary according to the insect, to the type of crop under investigation, to the number and size of samples possible, and to the importance of damage to the crop.

The problem investigated was that of estimating the total number of insects in a *single* field. This problem differs to some extent from that of estimating the population in an area such as a county.

Review of literature

A considerable amount of investigation on the best method for sampling in agronomic work has been carried out. Some of that work, pertinent to the present problem, is discussed below with two investigations on the technique of sampling for insects.

The usage of Wishart & Clapham (1929) may first be noted. To these workers "units" are "the ultimate parts of a sample", that is, the smallest area from which yield has been examined; "sampling-units" are the "parts of a sample which are located independently and at random within the area to be sampled. Each may consist of one or many units"; a "sample" is "the aggregate of sampling-units taken from the area".

Clapham (1929) made a study of various methods of sampling cereals from a plot. This work showed, first, systematic arrangement of sampling-units to give an invalid estimate of chance variability and so random drawing to be necessary, secondly, the variability of estimates to be much smaller when samples were drawn from within subplots than when drawn from the plot in general, and thirdly, drawing throughout the plot to be superior to drawing

from randomly chosen rows. The latter procedure was the least laborious and valid but gave a high error to estimates. Later, Clapham (1931) discussed the practical technique of locating sampling-units.

Clapham (1929) pointed out that a systematic arrangement of units may be combined into a sampling-unit, and Wishart & Clapham (1929) considered and employed sampling-units of complex patterns of units. Kalamkar (1932) made a uniformity trial on wheat, with the unit employed a half-metre of drill. He formed sampling-units in various ways from groups of four units and concluded that the only satisfactory sampling-unit is a strip running transversely to the direction of the rows.

Influenced by agronomic work of the type discussed above, Marshall (1936) made a study of the most suitable method of sampling a field in the determination of oviposition by the moth, *Heliothis obsoleta* Fabr. As had been found in agronomic work, so in this work Marshall found variations in eggs per 3 yd. of row to be greater between than within rows. He found the part, ascribable to sampling errors, of the variability between plots to be small with even 1 or 2 % sampling. A second study on sampling for insects is that of Fleming & Baker (1936). They made counts on numbers of larvae of *Popillia japonica* Newman present per unit area of 1 sq. ft. over four fairly large blocks of land and they recommended that a sample of at least 1 % be taken.

DESCRIPTION OF EXPERIMENTAL MATERIAL

Suitable material upon which to test theoretical results in the problem of sampling insect populations was found in the adult Colorado potato beetle, *Leptinotarsa decemlineata* Say. This insect is easily counted since it is both seen and collected rapidly. Such a count on the number of beetles present in a field near Chatham, Ontario, was made on 14 August 1935. This field was infested to the unusual extent of about two beetles to the linear foot of potato row. The field, a little more than an acre in extent, was fifty-eight rows of potatoes wide and about as broad as long. The plants were, on the average, spaced within the row a little more than a foot apart. The plot chosen for examination was forty-eight rows, or 124 ft. wide, and 96 ft. long. This plot included one margin of the field. The field was surrounded by various other crops.

THEORETICAL BASIS OF WORK

The paper of Neyman (1934) was the theoretical basis of the present work. Neyman discussed the general theory of sampling from strata. By the term strata, so far as the present work is concerned, is meant arbitrary subdivisions of an area of which the population is to be estimated. From within each stratum a certain number of sampling-units was selected. These sampling-units were of various kinds formed by combinations of a number of smaller basic units of

a fixed size. The terms sampling-unit and unit have been employed by Wishart & Clapham (1929) as previously discussed.

In much work, such as surveys, it will be desirable to make strata correspond at least roughly with obvious features, such as slope or wetness of land, which will affect the abundance of insects. In experimental work within one field, however, it is common practice to select areas that appear to be as nearly as possible homogeneous. Further, to have a uniform series of subareas may be practically convenient in making counts with a group of workers. Accordingly, equal strata will be commonly employed, and in the present paper the discussion was restricted to such strata. If each stratum is of the same area, that is, contains the same number of sampling-units, the equations and the numerical calculations involved in making estimates of population values are more simple than the general equations and calculations of Neyman (1934).

Denote by N the number of strata and by M the size of a stratum in terms of the number of sampling-units contained. M will, of course, vary with the size of the sampling-unit. Whatever the number be, each stratum will be divided into M sampling-units such that each is a potential sample.

Consider the notation for the total sampled population. Let X represent the total number of an insect in the area to be examined. Let u_{ij} denote the number of the insect in the jth sampling-unit ($j=1, 2, \ldots, M$) of the ith stratum ($i=1, 2, \ldots, N$) and $u_{..}$ denote the average number, calculated over the whole field, of the insect per sampling-unit. Then

$$X = NMu_{..}. \qquad \ldots\ldots(1)$$

Within the ith stratum let the mean value of u_{ij} be $u_{i.}$, and the variance

$$\sigma_i^2 = \frac{1}{M-1}\sum_{j=1}^{M}(u_{ij}-u_{i.})^2.$$

It should be noted that σ_i^2, as here defined, is $M/(M-1)$ times greater than the parallel quantity employed by Neyman (1934). For this discrepancy, allowance was made in all equations quoted.

Consider now the notation for the samples. Denote by m_i any number of sampling-units drawn from the ith stratum. Let the numbers of an insect found in the sampling-units of the ith stratum be $x_{i1}, x_{i2}, \ldots, x_{im_i}$ with mean $x_{i.}$, and with estimated variance,

$$s_i^2 = \frac{1}{m_i-1}\sum_{j=1}^{m_i}(x_{ij}-x_{i.})^2.$$

The best linear estimate of X, that is the estimate with minimum S.D., will be

$$F = M\sum_{i=1}^{N} x_{i.}. \qquad \ldots\ldots(2)$$

The standard deviation of F, when the m_i sampling-units have been drawn randomly, will be, following Neyman (1934),

$$\sigma_F = \sqrt{\left\{\sum_{i=1}^{N}\left(\frac{M_i(M_i-m_i)}{m_i}\sigma_i^2\right)\right\}}, \qquad \ldots\ldots(3)$$

where M_i is the total number of sampling-units in the ith stratum. Since, in the present work, all values of $M_i = M$, equation (3) reduces, so that

$$\sigma_F = \sqrt{\left\{M \sum_{i=1}^{N} \left(\frac{M-m_i}{m_i}\sigma_i^2\right)\right\}}. \qquad \text{......(4)}$$

A common system of apportioning sampling-units is to make the number from each stratum proportional to the magnitude of the stratum. In the present work the number of sampling-units drawn from each stratum would be the same, that is $m_i = m$ in all cases. Under these circumstances, equation (4) reduces, so that

$$\sigma_F = \sqrt{\left(\frac{M(M-m)}{m} \sum_{i=1}^{N} \sigma_i^2\right)}. \qquad \text{......(5)}$$

FORM IN WHICH DATA WERE COLLECTED AND ANALYSED

On the basis of the foregoing theoretical discussion the form of collection and of analysis of data on the number of beetles present in the observational area was determined. This area was divided into small, approximately square, units. The population of beetles in each unit was recorded. This count was the equivalent of a uniformity trial in agronomic work.

For the purpose of the present work a 2 ft. length of row was the unit of observation. To obtain these units, strings were run transversely to the rows of potatoes across the area at intervals of 2 ft. There were 2304 units involved. The numbor of beetles in each unit was counted.

Various types of sampling-unit were formed by combining adjacent units in various ways. For sampling-units of each of a number of given sizes, various shapes and orientations were examined. Compact sampling-units, not those compounded of scattered units, as suggested by Wishart & Clapham (1929), were employed. The compact form seemed the only one practically possible. In the course of the present work nine types of sampling-unit were investigated as listed, with reference numbers, in Table I. This table indicates for each type

TABLE I

The various types of sampling-unit employed

Sampling-unit	Width in units	Size $=k$	Orientation of long axis with respect to direction of rows
1	1	1	—
2	1	2	Parallel
3	1	2	Transverse
4	1	4	Parallel
5	1	4	Transverse
6	2	4	—
7	1	12	Parallel
8	1	12	Transverse
9	3	12	Transverse

of sampling-unit, first, the smallest dimension in terms of units, secondly, the number, k, of units embraced, and thirdly, the direction of the long axis with respect to the rows of potatoes.

The types of sampling-unit listed in Table I are shown diagrammatically in Fig. 1. On each form the number of the type is shown.

Having obtained the number of beetles in each unit, it became possible to determine by trial the best shape and orientation and also the best size for sampling-units. It may be expected that the occurrence of insects noted in a field will vary with the direction in which an observer moves in the field, and that of two directions at right angles, one will show more differentiation than the other.

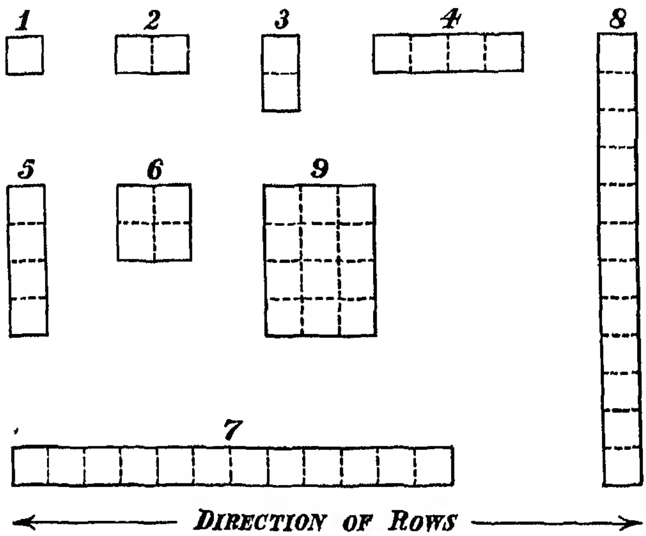


Fig. 1. The various types of sampling-unit employed.

Thus, in entomological work, Marshall (1936) found variability between rows to be greater than that within rows. Direction of ploughing and slope of a field also tend to differentiate the observations in certain directions. One may expect the population of phytophagous insects to be influenced by variability in the plants of a field. Also differences of shade and of wind in a field, migration along rows of plants, and point of ingress to a field, are all factors that tend to make the insect population variable in certain directions. Accordingly, the shape of soil surface forming a sampling-unit may be expected to be of importance in the determination of the accuracy of estimates made from a sample. Long narrow sampling-units running in the direction of greater differentiation should be the most efficient.

Just as various types of sampling-units were tested on the data collected, so might one have tested various types of strata. Presumably, the best type would be a long rectangle running in the direction of lesser variability. However, the

problem of type of stratum was not investigated, since, as is discussed below, it was thought advisable to fix the strata coterminal with the areas examined by each man.

It was necessary to cover a considerable area and to cover it in one day, since the population of beetles was changing rapidly from day to day. Accordingly, four men, A, B, C and D made counts. To each man were allotted four subareas, or strata, twelve units square. When the counting was arranged, the square form, in units, was chosen for the subareas assigned to each man, in case these subareas should have to serve as strata, because, within square strata sampling-units could be formed to the same extent longitudinally as transversely. The men were arranged in Latin square form so that personal effects should not be confused with trends across the field and so that the effect of each man might be discernible. The positions of the four men involved, and their collections are shown in Fig. 2.

D 1127	*B* 1331	*A* 628	*C* 430
C 658	*A* 635	*D* 969	*B* 758
B 869	*D* 794	*C* 560	*A* 411
A 523	*C* 490	*B* 213	*D* 517

← Direction of rows →

Fig. 2. The total numbers of beetles taken in each subarea assigned to four men.

PRESENTATION OF DATA

The primary data upon which the present paper was based are given completely in Table VI of the Appendix. In Fig. 3 the general nature of the variation in population, throughout the area studied, is indicated. In this figure the population density over the area examined is indicated by the population on 144 equal constituent subareas, four units square. The counts for the 16 units in each subarea were totalled. In the figure each subarea is represented by a black spot of which the area is proportional to the number of beetles found on the subarea.

HOMOGENEITY OF DATA

Before considering the questions, indicated in the foregoing discussion, of goodness of various sampling-units, or of the efficiency of the various methods of apportioning sampling-units, the general nature of the insect distribution in

the observational area was investigated. The data of Table VI, which are presented graphically in Fig. 3, suggested there to have been much variation from stratum to stratum in the number of insects present. The part of variability ascribable to differences between the observers and also the magnitude of the chance variability between the strata, as compared with that within strata, were considered.

For the 2304 sampling-units of type no. 1, with $k=1$, the total variability was broken into a part within strata and a part between strata. Since the four men, who made counts, were assigned in the manner of a Latin square, as shown in Fig. 2, the variability between strata was broken into parts ascribable to

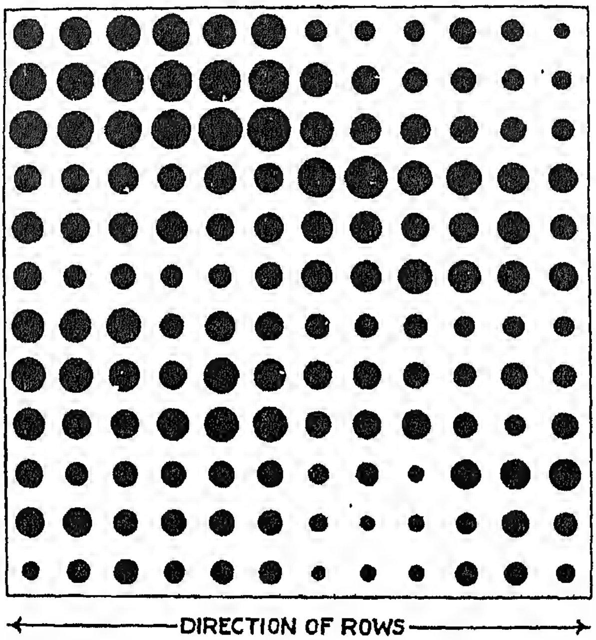


Fig. 3. Diagrammatic representation of population density over the area under observation.

rows, columns, men and a remainder term. The analysis is presented below, although subsequent work by the present writer to be published later has shown that analysis of the type carried out, involving the use of normal theory, is not strictly applicable to entomological data because the chance variability of the number of insects observed is related to that number.

In the following analysis of variance, the mean square from within strata can be compared with the mean square from the remainder for between strata. This remainder is free from differences ascribable to the rows, columns and men. One is, then, comparing the chance variability for small subareas within a small total area with the chance variability for larger subareas within a larger total area. The very great difference observed in the following tabulation between

these two variabilities showed that the strata, apart from the differences introduced by observers and even when row and column effects were removed, varied much more than sampling-units within strata. Accordingly, a very great amount of the variability within the area should be controllable by stratification.

Variability ascribed to	Degrees of freedom	Sum of squares	Mean square
Rows	3	1,695·8	565·3
Columns	3	2,925·8	975·3
Men	3	2,235·0	745·0
Remainder between strata	6	1,627·7	271·3
Within strata	2288	26,064·7	11·4
Total	2303	34,549·1	

In the analysis of variance, the remainder variability between strata, free from the variability of rows, columns and men, consisted of the chance variability between strata and possibly, also, of a differential response by a given man in the various strata in which he worked. To assess the significance of the variability of men the appropriate sum of squares must be referred to this remainder term, since the differences between men are subject to the chance variability between strata.

When the mean square for the men was compared with the mean square for the remainder, the result was within the 0·05 level of probability, so that the differences between men were not proved significant. The effect of the men was not appreciable over the variability from stratum to stratum, possibly, because this variability was estimated with only 6 degrees of freedom. There is further evidence, however, in the primary data of Table VI, which leads one to judge that the effect of the men was appreciable. In these data the counts made by each man may be viewed in either dimension as 12 rows of units. Consideration of such rows suggested that the differences between adjacent rows in the area covered by one man were smaller than the differences between adjacent rows on the borders of the areas covered by two men. Such differences are free from the great chance variability of strata. In considering these differences the uniformity from observer to observer of the work done can be studied without supposing that any one man worked with uniform efficiency. One can show statistically that the men collected differently.

Since the differential efficiency with which the men collected would make the strata composed of units collected by more than one man unduly heterogeneous, it was thought advisable to use as strata the subareas worked over by each man. In this procedure the variability introduced by the men was combined with regional variability.

Any man must miss insects when he is making a count and, from the discussion above, the number missed appears to vary with the observer. These considerations modify the meaning of X, which must be regarded as the total number of insects that may be found, with complete examination, by the particular men employed in counting, rather than the total number of insects present in the area. In field work where one is making counts on one area to be compared with those from another area, variability in the performance of observers would occur and must be taken into consideration. Thus, in experimental work all counting on a given block may be done by a single man or in counts for a survey of a number of fields, each man may do a constant fraction of the work in each field.

Efficiency of various types of sampling-unit

The relative efficiency of various types of sampling-unit of the same magnitude, in the case where m sampling-units were drawn from each stratum, was judged by means of the following equation, derived from equation (5) by putting $M_0 = MN$, $m_0 = mN$:

$$\sigma_F = \sqrt{\left(\frac{(M_0 - m_0) M_0}{m_0} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2\right)}. \qquad \text{......(6)}$$

It can be seen that, when the total number, M_0, of sampling-units in the field, the total number, m_0, of sampling-units to be drawn, and the number, N, of strata, are fixed, then the accuracy of the estimate, F, is determined by the average variance within strata, $\sigma_0^2 = \sum_{i=1}^{N} \sigma_i^2/N$. Accordingly, among several forms of sampling-unit, that which gives the smallest value to σ_0^2 gives the greatest accuracy to the estimate, F.

The relative efficiency of sampling-units differing in magnitude was judged by means of equation (7), shown below. When the number of strata and the fraction of the area to be sampled are fixed, σ_F may be supposed affected by increase in size of the sampling-unit. Suppose that when the sampling-unit is of unit size one obtains σ'_F, and, when the sampling-unit consists of $k > 1$ units, one obtains σ''_F. When $k = 1$ let M have the value M' and m the value m'. For each value of k there will be a value for σ_0^2 and values, M'' and m'', for M and m, such that $\dfrac{M''}{M'} = \dfrac{m''}{m'} = \dfrac{1}{k}$. From equation (5),

$$\frac{\sigma''_F}{\sigma'_F} = \sqrt{\frac{\sigma_0''^2}{k}} \div \sqrt{\sigma_0'^2}. \qquad \text{......(7)}$$

From equation (7) it is apparent that for any sampling-unit, of $k > 1$ units, the relative magnitude of σ''_F and of σ'_F will vary as the relative magnitude of $\sqrt{(\sigma_0''^2/k)}$ and of $\sqrt{\sigma_0'^2}$. Accordingly, the relative efficiency of sampling-units of any sizes may be judged by the relative magnitudes of $\sigma_0''^2/k$. The expectations of $\sigma_0''^2/k$ would, of course, be the same if the insects involved were distributed quite

randomly over the strata. Under these circumstances, large or small sampling-units would be equally good.

For sampling-units of various forms and of various sizes, that is with various values of k, $\sigma_0''^2$ and $\sigma_0''^2/k$ are shown in Table II.

TABLE II

Mean variance, within the sixteen strata, for various sampling-units

Sampling-unit type of Table I	k	$\sigma_0''^2$	$\sigma_0''^2/k$
1	1	11·39	11·39
2	2	25·27	12·63
3	2	24·50	12·25
4	4	61·39	15·35
5	4	50·12	12·53
6	4	58·08	14·52
7	12	298·91	24·91
8	12	150·37	12·53
9	12	233·76	19·48

From the values of $\sigma_0''^2$ it can be seen that, within each size class of sampling-units, the long narrow form (nos. 3, 5 and 8) running transversely to the direction of the rows was the most efficient, and that the long narrow form (nos. 2, 4 and 7) running in the direction of the rows was the least efficient. Such a result is explained by the apparent correlation of the number of beetles on units along the rows of potatoes, as shown in Table VI. It is of interest to note that for long narrow sampling-units running transversely to the rows the value of $\sigma_0''^2/k$ changed but little. Such a result means that in the direction considered, the population of the insect was practically randomly distributed. Bearing in mind equation (7), it can be seen from the values, $\sigma_0''^2/k$, that the value of σ_F'' was in general greater as the value of k, or the size of the sampling-unit, increased.

The values shown in Table II indicated that the estimate of F from a given amount of sampling had least variability with the smallest sampling-unit employed. While this conclusion applies to the case where m_i was the same for all strata, a similar effect probably occurs in the more general case, where m_i differs from stratum to stratum.

Day (1920) pointed out that long plats are only best when the length of the plat lies along the direction of the greater changes of soil fertility. He suggested that if the direction of greater differentiation is unknown square plats are probably best. From the data of the present work, even when narrow sampling-units running transversely to the direction of the rows were employed, the estimate of F was a little less reliable for $k > 1$ than for $k = 1$. In practice, since it may happen that some phenomenon such as slope acts against the effect of row direction

in deciding the direction of greater variability, one would not necessarily know the direction in which long sampling-units should run to get the best results. Accordingly, it is probable that, in general, the best results would be obtained with the smallest sampling-unit.

Employment of stratification

When the question of the type of sampling-unit is decided, the problem of the best method for apportioning sampling-units must be considered. Accordingly, for the data of the present work, the percentage of area, i.e. $100m_0/M_0$, which would need to have been sampled to secure a specified degree of accuracy was calculated. The degree of accuracy was expressed in the familiar form of standard deviation of the estimate of population in terms of the population, i.e. by σ_F/X. For sampling-units of a given type there was found the total number, m_0, (*a*) necessary in order to obtain a given value of σ_F without stratification, and (*b*) necessary with the number, m_i, examined in each stratum proportional to M_i. The respective values, m_0, were found simply from equation (5), for in the case of no stratification, $N=1$. The value of m_0 was, also, determined for m_i made proportional to σ_i, as will be discussed in the next section of this paper. Table III indicates the proportional amount of sampling necessary to ensure values of σ_F/X equal to 0·01 and 0·10, by employing the various methods of apportioning sampling-units of various sizes previously discussed. The calculations were made for sampling-units of size $k=1$, 2, 4 and 12, when the best shaped and orientated sampling-unit, i.e. nos. 1, 3, 5 and 8 of Table I, in each size-class was employed.

From Table III it is apparent that there was a considerable reduction in the percentage of the area to be covered when stratification was employed. The reduction was greatest when the sampling-units were large and also when the desired degree of accuracy was low. The results also indicate that further reduction was effected when the number of sampling-units apportioned to each stratum was proportional to the standard deviation per stratum. It will be noted that such a system is hardly practicable unless k, the size of sampling-units, is of such a value that M, the total number of sampling-units per stratum, is great.

Optimal apportionment of the work of sampling

If stratification be employed, the value of σ_F is not reduced to the lowest level possible for a given amount of sampling by making the values of m_i proportional to M_i. Neyman (1934) considered how $m_0=\sum_{i=1}^{N} m_i$ sampling-units should be apportioned to the N strata so that σ_F shall be minimal. He found that σ_F^2 is minimal if the values, m_i, are proportional to $M_i\sigma_i$, and then

$$\sigma_F^2 = \frac{M_0-m_0}{m_0}\sum_{i=1}^{N}(M_i\sigma_i^2) - \frac{M_0}{m_0}\sum_{i=1}^{N} M_i\left\{\sigma_i - \frac{1}{M_0}\sum_{i=1}^{N}(M_i\sigma_i)^2\right\}, \quad \ldots\ldots(8)$$

where $M_0 = MN$. In the present work, where M was the same for each stratum, σ_F^2 was minimal if the values, m_i, were proportional to σ_i, and equation (8) reduces so that

$$\sigma_F = \sqrt{\left\{\frac{M^2}{m_0}\left(\sum_{i=1}^{N} \sigma_i\right)^2 - M\sum_{i=1}^{N} \sigma_i^2\right\}}. \qquad \text{......(9)}$$

TABLE III

The percentage of area which must be sampled in order to obtain a specified degree of accuracy

Sampling-unit type of Table I	k	Without stratification	With stratification: m_i proportional to M_i	With stratification: m_i proportional to σ_i
		Degree of accuracy, $\sigma_F/X = 0{\cdot}01$		
1	1	74·37	68·79	61·72†
3	2	79·01	70·32	
5	4	83·91	70·80	
8	12	91·55	70·80	
		Degree of accuracy, $\sigma_F/X = 0{\cdot}10$		
1	1	2·82	2·16	1·95†
3	2	3·63	2·31*	
5	4	4·96	2·37*	
8	12	9·77	2·37*	

* Note that actually it would have been impossible to use these values since the necessary minimum of two sampling-units per stratum would not have been attained.

† These solutions were somewhat unreal, since, with the levels of sampling and with the variability per stratum involved: (1) in the case of $\sigma_F/X = 0{\cdot}01$, although there were only 144 sampling-units, 150 would have to have been apportioned the most variable stratum; (2) in the case of $\sigma_F/X = 0{\cdot}10$, only 1·17 sampling-units would have to have been apportioned the least variable stratum. The last column is discussed at length in the next section of this paper.

To complete the discussion on Table III it may be noted that values of m_0, necessary to obtain a given value of σ_F, as shown in that table, can be found simply from equation (9) when m_i is made proportional to σ_i. In computing these values of m_0, the requisite integrality of the number of sampling-units per stratum was ignored and the limit of accuracy possible by this method of sampling was found. This apportionment was made with only sampling-unit no. 1 ($k = 1$), since for larger sampling-units the results tend to be meaningless with the present data. For example, in the case of $k = 12$, m_i in the least variable stratum could not fall below 2 and in the most variable stratum could not exceed 12. The largest value of σ_i is 4·01 times greater than the smallest, so practically only one level of such sampling was possible. As can be seen in Table III, even for $k = 1$, with the range of accuracy considered, one or two of the assignments to strata with extreme variability were unreal.

Approach to optimal apportionment in practice

In the foregoing discussion it has been pointed out that, if m_0 sampling-units are apportioned to the strata so that m_i is proportional to σ_i, σ_F is minimal. Since in practice one would not know the values, σ_i, an optimal apportionment of sampling could not be made exactly. Estimates, s_i, made from preliminary sampling, might be employed, however, in place of the true values, σ_i.

One can find the probability that, when m_i is made proportional to s_i, the value of σ_F will be smaller than when m_i is constant. It can be seen that, by the first system of apportionment, σ_F will be subject to chance variation depending upon the estimates, s_i. One can determine, however, for a preliminary sample of any size, the probability that σ_F will be greater under the first system than under the second. This determination can be made by using the moments of $z = m_0\sigma_F^2/M^2$ as given approximately by Sukhatme (1935).

The probability that the value of σ_F would be less with m_i proportional to s_i, based on 15, 10 or 6 sampling-units, than with m_i constant, was found by applying the procedure of Sukhatme to the data of the present paper. For each number, 15, 10 or 6, the first three moments of z were found. With preliminary samples of 15 and 10, β_1 was 1·38 and 2·01, respectively, so that a type III Pearson curve was fitted by the first three moments. In the case of a preliminary sample of 6, $\beta_1 = 5{\cdot}66$ was so great that a type III curve was fitted by the first two moments and the start, which comes from the value of σ_F when m_i is proportional to σ_i. From these three curves the probability that the value of σ_F would be less with m_i proportional to s_i than with m_i constant was 0·99973, 0·957 and 0·519, respectively. These probabilities show that improvement in the estimate, F, is almost certain to result from a preliminary sample of 15, will probably result from one of 10 but doubtfully so from one of 6. It should be noted that Sukhatme did not advise using a preliminary sample smaller than 15.

Sukhatme suggested, as illustrated in the next section, that one might incorporate preliminary sampling, made to estimate s_i, with supplementary sampling in a total sample to be used in forming the required estimate of population.

It is conceivable that a preliminary estimate of the relative magnitude of the values σ_i might be made from a cursory or visual survey rather than from exact preliminary sampling. Although it is difficult to appreciate variability, advantage might be taken of the relationship that exists between the number of insects per unit area and the chance variability of that number, since the level of population is easily appreciated. Thus if a field man were to judge from a visual survey that an insect were four times more numerous in one stratum than in a second, then, since the standard deviation should vary approximately as the root of the mean number of insects per sampling-unit, the first stratum should be sampled twice as heavily as the second. Whether an efficient apportionment could be made on such a basis would have to be tested in practice.

APPLICATION OF RESULTS

In order to illustrate the application of the foregoing work it is supposed that one wish to make an estimate of the number of beetles in the area considered in the present paper. The practical procedure to be followed is indicated below.

It is necessary in the first place to fix the type of sampling-unit to be employed. For the present illustration it is supposed that one choose type no. 1 of Table I. One must fix randomly the position within the strata or field of the sampling-units to be examined. The choice involved may be made by using random sampling numbers, Tippett (1927).

In the second place it is necessary to fix the fraction of the area to be sampled, that is to choose a value of m_0 in relation to M_0. In the present illustration two cases are considered, the first where 25 % of the total area is comprised in the sample, and the second where approximately 15 % is comprised.

In the third place one must decide how the work of sampling shall be apportioned. This work can be done without stratification and with stratification. With stratification it can be done with m_i constant for all strata and also with m_i approximately proportional to σ_i.

Whatever method of drawing sampling-units be employed, the total number is 576, if 25 % of the total area is to be examined. In the case where no stratification is employed the sampling-units may simply be drawn successively and independently. The procedure is equally simple if 36 sampling-units are drawn from each of the 16 strata previously discussed. In both cases, F, the estimate of total population, may be made from equation (2). If approximately 15 % of the total area is to be examined, then with no stratification 336 sampling-units must be chosen or with stratification 21 sampling-units must be examined in each stratum. The case, however, where an attempt is made to secure values of m_i, approximately proportional to σ_i, requires more detailed discussion. From this detailed discussion the procedure for the more simple cases will be obvious.

It is supposed that in the present population study, preliminary and supplementary samplings are possible. In order to make a well apportioned sample of 25 % a preliminary sample of 15 sampling-units per stratum is made first. For an illustration consider the procedure in the first stratum, for which the counts are shown in the upper left-hand corner of Table VI. The position of any sampling-unit may be represented by one number indicating the column and another the row in which it lies. In such terms, fifteen positions, drawn randomly without replacement, are: 2–3, 2–5, 2–8, 3–4, 3–5, 4–3, 4–8, 6–12, 7–6, 8–8, 8–11, 9–7, 9–8, 11–9, 12–8. Examining the sampling-units indicated by these numbers one obtains the fifteen observations: 9, 5, 3, 7, 7, 8, 7, 8, 5, 14, 1, 7, 11, 10, 4. From these observations one calculates $s_1 = 3{\cdot}26$, as shown in Table IV. In that table there is shown for each stratum a value, s_i. The order

in which the strata are listed is down the columns taken from left to right in Table VI.

It is now necessary to find values of m_i proportional to s_i so that $\sum_{i=1}^{16} m_i = 576$, since a 25 % sample is to be taken from the 2304 sampling-units in the whole area. Thus for the first stratum, $s_1 = 3{\cdot}26$ and, since $\sum_{i=1}^{16} s_i = 47{\cdot}87$,

$$m_i = (3{\cdot}26/47{\cdot}87)\, 576 = 39,$$

as shown in Table IV. In that table there is shown for each stratum a value, m_i.

In order to make a sample of 15 %, a preliminary sample of 6 sampling-units per stratum is made first. The procedure is similar to that shown above for a sample of 25 % with a preliminary survey of 15 sampling-units. The values, s_i, and the corresponding values of m_i are shown in Table IV.

TABLE IV

Numerical results in the process of sampling with m_i approximately proportional to σ_i

Stratum no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
σ_i	5·39	2·77	3·48	3·21	5·36	2·78	3 76	2 29	3·27	4·57	2·35	1·34	2·67	3·14	2·19	2·59
	Preliminary sample of 15 for a sample of 25 %															
s_i	3·26	3·00	2·93	3·45	5·26	2·78	3·92	2·81	2·97	4·19	2·28	1·68	2·89	2·08	1·74	2·63
m_i	39	36	35	42	63	34	47	34	36	50	27	20	35	25	21	32
$\sum_{j=1}^{m_i} x_{ij}$	284	177	215	180	597	156	272	117	164	332	116	42	115	136	49	126
	Preliminary sample of 6 for a sample of 15 %															
s_i	5·22	1·94	3·95	2·68	5·01	2·93	3·94	0·55	1·03	3·44	1·75	1·17	1·67	1·79	2·61	2·79
m_i	41	16	31	21	40	23	31	5	8	27	14	9	13	14	21	22
$\sum_{j=1}^{m_i} x_{ij}$	337	59	177	70	361	104	150	23	43	147	42	15	42	66	73	97

The preliminary drawings must be supplemented to make m_i as great in each case as is required in Table IV. Thus, in the case of the first stratum when a sample of 25 % is desired, $m_i = 39$. Accordingly, it is necessary to make a supplementary sample of 24 sampling-units, and the previous random drawing without replacement must be continued. Twenty-four such sampling-units, in the terms previously employed, are: 1–1, 1–2, 2–7, 2–11, 3–1, 3–2, 3–3, 4–5, 4–11, 5–4, 5–11, 6–2, 6–5, 6–6, 6–8, 7–11, 8–2, 8–7, 9–9, 10–9, 11–2, 11–4, 12–5, 12–11. By reference to Table VI, the observations corresponding to these positions can be discovered, thus, one obtains: 2, 0, 10, 12, etc. Over all the strata 336 supplementary sampling-units must be found and then, from equation (2), F can

be calculated on the basis of 576 sampling-units. In finding F, one must estimate the mean for each stratum; for instance, find that

$$x_1' = \left(\sum_{j=1}^{m_1} x_{1j}\right) \Big/ m_1 = 284/39 = 7{\cdot}28.$$

From the values of $\sum_{j=1}^{m_i} x_{ij}$ in Table IV, it can be seen that

$$F = 144(284/39 + 177/36 + 215/35 + \ldots + 126/32) = 11{,}305.$$

When a sample of 15 % is desired, 241 supplementary drawings must be made over all the strata so that F can be calculated on the basis of 336 sampling-units.

It may be of interest to note with what accuracy X would be estimated by each of the three methods, first, sampling without stratification, second, uniform sampling with stratification and third, sampling within strata in proportion to the values of s_i in Table IV. Accordingly, σ_F for each method and also the minimal value of σ_F, which is obtained when m_i is proportional to σ_i, are shown in Table V. It should be noted that σ_F has a fixed value for a fixed amount of sampling in the first two cases, but in the third case the value of σ_F depends upon the particular values of m_i shown in Table IV and so, if fresh values of s_i were calculated, the values of m_i and of σ_F would probably differ from those in the third column of Table V.

TABLE V

The value of σ_F obtained from various methods of sampling

	Without stratification	Stratification with m_i constant	Stratification with m_i proportional to s_i	Stratification with m_i proportional to σ_i
Preliminary sample of 15 for a sample of 25%	322·0	280·6	268·1	260·7
Preliminary sample of 6 for a sample of 15%	449·9	392·1	403·4	367·8

From the values of σ_F in Table V it can be seen that a considerable improvement in the estimate, F, was obtained by stratification. In the attempt to secure further improvement by making m_i proportional to s_i, rather than constant, the results are such as would be anticipated from the discussion of the previous section for, on the one hand, when the preliminary survey consists of 15 sampling-units, an improvement occurs, but, on the other hand, when only of 6 sampling-units, the results are not as good as those from general uniform sampling. The final column in the table shows the limit of accuracy that would be attained were the values of σ_i known. In the case of the preliminary samples of 15 the guesses at m_i obtained from s_i have been good enough to make the value of σ_F approach fairly closely to its ideal minimum of 260·7.

Summary and conclusions

The discussion of Neyman (1934), on making estimates of population, has been applied to the entomological problem of finding the number of the Colorado potato beetle, *Leptinotarsa decemlineata* Say, on a heavily infested field. Observations were made on the population of beetles per 2 ft. of row of potatoes for entire rows in the area considered. From these observations, sampling-units were variously formed and their relative desirability studied.

Neyman's general equations are simplified and computation lightened when the strata, or subdivisions of the field, can be made equal. In the field considered, it was found that the variability was much greater between strata than within strata. Although this variability was due to some extent to differences in the work of the different observers, for the purpose of discussion the entire variability was regarded as real. It was found that a marked reduction in the area which must be examined to secure a given degree of accuracy in the estimate of population could be secured, first, by stratification and secondly, by making the number of sampling-units examined in a given stratum proportional to the standard deviation for the sampling-units in that stratum. These standard deviations will not be, in general, known, and the experimental data have been used to illustrate how their values may be replaced by estimates obtained from a preliminary survey, on the lines suggested by Sukhatme (1935).

In the present case it was found that, if the total sampling were to amount to 25 % of the whole, then a preliminary survey involving the selection of 15 sampling-units per stratum (10·4 % of the whole) would have led to a definite reduction in the standard error of estimate (as shown in Table V). Were the total sampling to amount to only 15 % of the whole, a preliminary survey of 6 sampling-units per stratum (4·2 % of the whole) was found to be inadequate.

The main object of this paper has been to investigate, from a statistical point of view, the consequences of applying certain sampling methods to an insect population. The question of how far, in following the Neyman-Sukhatme method, any extra inconvenience due to unequal numbers of sampling units per stratum or the need for preliminary sampling, would be justified in practice by the extra accuracy gained, is of course a matter requiring fuller consideration by the entomologist.

Acknowledgements

This study is based upon suggestions made by Prof. J. Neyman, to whom the writer is deeply indebted. The writer acknowledges with gratitude the guidance and advice of Prof. E. S. Pearson in the execution of the work. Mr G. M. Stirrett and Mr D. A. Arnott of the Canadian Dominion Entomological Branch gave advice on field work, and Dr M. S. Bartlett made helpful criticisms regarding the final form of the paper.

REFERENCES

R. (1929). "The estimation of yield in cereal crops by sampling methods." . *Sci.* **19**, 214.

"Studies in sampling technique: cereal experiments. I. Field technique." *Sci.* **21**, 366.

'1920). "The relation of size, shape, and number of replications of plats to error in field experimentation." *J. Amer. Soc. Agron.* **12**, 100.

E. & BAKER, F. E. (1936). "A method for estimating populations of larvae panese beetle in the field." *J. Agric. Res.* **53**, 319.

R. J. (1932). "A study in sampling technique with wheat." *J. Agric. Sci.*

(1936). "The distribution and sampling of insect populations in the field ial reference to the American bollworm, *Heliothis obsoleta* Fabr." *Ann. Appl.* 133.

1934). "On the two different aspects of the representative method." *J.R.* c. **97**, 558.

V. (1935). "Contribution to the theory of the representative method." *st. Soc.* Suppl. **2**, 253.

. C. (1927). *Random Sampling Numbers.* Tracts for Computers, XV. Camb. ss.

, CLAPHAM, A. R. (1929). "A study in sampling technique: the effect of rtilizers on the yield of potatoes." *J Agric. Sci.* **19**, 600.

APPENDIX ON PRIMARY DATA

er of beetles per unit, over the area examined, is shown in Table VI. discussed, the area was broken into sixteen subareas, twelve units nits of these subareas are indicated by straight lines.

REFERENCES

CLAPHAM, A. R. (1929). "The estimation of yield in cereal crops by sampling methods." *J. Agric. Sci.* **19**, 214.

—— (1931). "Studies in sampling technique: cereal experiments. I. Field technique." *J. Agric. Sci.* **21**, 366.

DAY, J. W. (1920). "The relation of size, shape, and number of replications of plats to probable error in field experimentation." *J. Amer. Soc. Agron.* **12**, 100.

FLEMING, W. E. & BAKER, F. E. (1936). "A method for estimating populations of larvae of the Japanese beetle in the field." *J. Agric. Res.* **53**, 319.

KALAMKAR, R. J. (1932). "A study in sampling technique with wheat." *J. Agric. Sci.* **22**, 783.

MARSHALL, J. (1936). "The distribution and sampling of insect populations in the field with special reference to the American bollworm, *Heliothis obsoleta* Fabr." *Ann. Appl. Biol.* **23**, 133.

NEYMAN, J. (1934). "On the two different aspects of the representative method." *J.R. Statist. Soc.* **97**, 558.

SUKHATME, P. V. (1935). "Contribution to the theory of the representative method." *J.R. Statist. Soc.* Suppl. **2**, 253.

TIPPETT, L. H. C. (1927). *Random Sampling Numbers.* Tracts for Computers, XV. Camb. Univ. Press.

WISHART, J. & CLAPHAM, A. R. (1929). "A study in sampling technique: the effect of artificial fertilizers on the yield of potatoes." *J. Agric. Sci.* **19**, 600.

APPENDIX ON PRIMARY DATA

The number of beetles per unit, over the area examined, is shown in Table VI. As previously discussed, the area was broken into sixteen subareas, twelve units square. The limits of these subareas are indicated by straight lines.

THE COMPARATIVE ADVANTAGES OF SYSTEMATIC AND RANDOMIZED ARRANGEMENTS IN THE DESIGN OF AGRICULTURAL AND BIOLOGICAL EXPERIMENTS

By F. YATES

1. Introduction

Ever since the introduction of the principle of randomization into replicated experiments it has been realized that certain of the arrangements generated by the randomization process were likely to be less accurate than others; consequently there has always been a conflict whether the general efficacy of a set of experiments might not be improved by the rejection of those arrangements which appeared *a priori* less accurate. In its extreme form such a procedure results in the rejection of the principle of randomization altogether, and the selection of one or more arrangements which on account of their balance or other features especially appeal to the experimenter.

Those who habitually use random arrangements may have thought that the issue was finally settled in favour of randomization, but the recent recrudescence of the dispute in the scientific journals, and the continued use of systematic arrangements in agricultural field trials, make it clear that there is still a considerable body of opinion which favour such arrangements. A review of the recent arguments, and a re-examination of the numerical material cited in support of them, may therefore be of value.

2. Statistical principles

The statistical treatment of the results of replicated experiments is usually based on the assumption of the normal law of error, and the general structure of the analysis is derived by the method of least squares.

The method of least squares was originally developed by Gauss, for the purpose of deriving the best estimates of unknown quantities from observational material in astronomy and geodesy. Gosset's discovery of the t distribution, and Fisher's extension to the z distribution, have provided exact tests of significance when, as usually occurs in practice, the degrees of freedom for error are few. The introduction of the procedure of the analysis of variance by Fisher has also considerably facilitated the arithmetical computations, particularly in the type of results that arise from planned experiments.

These modern advances, in their turn, have led to the wider recognition in practical work of the many different sources of variation to which experimental and observational material is subject. Without exact tests of significance and

the technique of the analysis of variance the assessment of these various components of variation would be a difficult and involved business.

For its correct application the method of least squares requires that any components of variation which are not eliminated by the design shall be normally and *independently* distributed. Now it is immediately evident that the yields of agricultural field plots (even after allowing for the effects of local control, such as blocks) are not independently distributed. Neighbouring plots tend to be positively correlated. This destroys the whole theoretical basis of the method of least squares, and in particular is liable to vitiate completely the estimates of error and tests of significance.

The difficulty can be met, as Fisher perceived, by the introduction of randomization into the design. This has the effect of removing the disturbance due to the correlation of neighbouring plots, so that yields can be treated *as if* their errors were uncorrelated. Adequate randomization requires that if all possible arrangements generated by the randomization process are put down in turn on the same set of yields (such as those from a uniformity trial), then the average of the mean squares for the (dummy) treatments is equal to the average of the mean squares for error. If this condition is not fulfilled it can easily be shown that certain types of correlation in the original material will give rise to biases in the estimate of error; such biases, if they exist, cannot fail to disturb the ordinary tests of significance.

Systematic arrangements lack this necessary element of randomization, and consequently their analysis by the method of least squares can never have the same objective validity as has the similar analysis of random arrangements. It is sometimes contended that the latter analysis is not really valid because the original material is not normally distributed, or in some other way fails to satisfy the conditions required for analysis by least squares. Actually, however, it is known that the majority of material that the experimenter has to handle does fulfil the required conditions sufficiently nearly, provided that a proper process of randomization is adopted. Consequently this contention must be regarded rather as a debating point than as a serious objection which will in its turn justify the abandonment of randomization.

3. The advantages and disadvantages of systematic arrangements

The advantages claimed for systematic arrangements of the "balanced" type are that they give more accurate results than do random arrangements, and that they are more easy to carry out in the field.

The disadvantages are as follows:

(1) There can be no assurance that the estimate of error is unbiased, however this estimate is arrived at, and the objectivity of the tests of significance is consequently lost.

(2) Many different methods of estimating the error can reasonably be advocated, so that the tests of significance are not even unique.

(3) The comparisons of different pairs of treatments are subject to different errors, so that even if the estimate of error is reasonably unbiased, it cannot be used to test individual differences.

(4) Biases may be introduced into the treatment means, owing to the pattern of the systematic arrangement coinciding with some fertility pattern in the field, and this bias may persist over whole groups of experiments owing to the arrangement being the same in all. Competition between plots with different treatments which always fall next to one another may produce similar effects.

The first disadvantage is admitted by some, but by no means all, of the advocates of systematic arrangements. Coupled with the admission is usually the plea that fully unbiased estimates of the errors of single experiments are not really required. Thus Gosset (1937) has discussed the point at some length, but Neyman (1937) has presented results which purport to show that two tests of significance which have recently been proposed for the half-drill strip arrangement are substantially accurate.

Gosset also recognized the second disadvantage, but maintained that it applied only to the half-drill strip method, whereas he himself provided an example in the same paper of another type of systematic arrangement for which many different methods of estimating the error immediately suggest themselves.

The third disadvantage has, so far as I know, never been fully recognized by the advocates of systematic arrangements. Indeed it is sometimes claimed to be an actual advantage.

The fourth disadvantage has, of course, been recognized for a long time, but the advocates of systematic arrangements have always maintained that the danger (except possibly in rare instances) can be avoided by care and foresight on the part of the experimenter.

It is very difficult to refute or substantiate this last claim, since from the nature of the case any biases that are in fact introduced will not be recognized as such, being attributed to treatment effects. Clearly biases affecting a whole group of experiments can be avoided by re-randomizing the treatments in each experiment, though this results in some loss of simplicity in execution, and is by no means always done. It is worth noting, however, that randomization appeals to practical experimenters more because it eliminates biases in the treatment means than because it provides a valid estimate of error. It is only when faced with the task of reducing and co-ordinating the results of large numbers of experiments, and of increasing the efficiency of future experiments, that they fully appreciate the existence of such estimates.

In short, randomization provides an assurance, not only to the experimenter, but to others who may be more sceptical than he, that the magnitude of the ordinary sources of disturbance, other than those eliminated by the arrangement,

has been evaluated by means of the estimate of error. It does not, of course, provide a panacea which removes all need for care and foresight: it cannot take account of types of disturbance which act selectively on the various treatments or varieties (e.g. bird damage), and a badly planned or carelessly executed experiment will still be inaccurate even though it is randomized, but the experimenter will at least know of its inaccuracy.

In co-operative experiments carried out by a number of workers at different places, where close supervision is frequently both difficult and expensive, and many of the workers have little training in experimental work, this assurance is doubly valuable.

The real question at issue, therefore, is whether the gain in accuracy and simplicity of execution are of such magnitude that they outweigh the manifest disadvantages of systematic arrangements. The advocates of systematic arrangements have claimed that the gain in accuracy, at least in certain cases, is very considerable: thus Hudson, quoted by Gosset in an appendix to his paper (1937), gives a set of comparisons between random and systematic arrangements in which the random arrangements gave, on the average, only one half the information that was yielded by the systematic arrangements.

Hudson's investigation, however, cannot be regarded as sufficiently extensive, or sufficiently representative of ordinary experimental practice to provide a fair estimate of the gain in accuracy due to systematic arrangements. In practice the average gain will probably be found to be decidedly smaller. Indeed in a line of research in which the experimental technique is actively developing, the random arrangements actually used are likely to be *more* accurate than the accepted systematic arrangements, since the unequivocable information that is provided on the error of each experiment as it is conducted, itself leads to advances in technique which far outweigh the small gain that might theoretically result from the use of some especially favourable systematic arrangement.

In the subsequent sections of this paper the above points will be examined in more detail. In the next section the special points that arise in connexion with the half-drill strip method will be discussed. I have chosen this particular type of systematic arrangement, not because I consider it of special importance, but because it was Gosset's advocacy of this arrangement that gave rise to the recent controversy, and because it does provide an excellent example of the many defects of even the simplest of systematic arrangements.

4. Barbacki and Fisher's test of the half-drill strip method

Following Gosset's advocacy of a new method of estimating the error of half-drill strip arrangements, with his general endorsement of the utility of this design (1936), Barbacki & Fisher (1936) imposed a half-drill strip arrangement on the yields of a uniformity trial on wheat, reported by Wiebe (1935).

Wiebe's trial consisted of 125 rows, harvested in 15 ft. lengths, twelve from

each row, each length being separated from the next in the row by a path. Barbacki & Fisher grouped these rows by sixes, omitting one row between each pair of groups so as to simulate a half-drill strip design. Thus from the first 104 rows they obtained eight pairs of half-drill strips, or four sandwiches ($ABBA$), for each 15 ft. length, i.e. forty-eight sandwiches in all. These forty-eight sandwiches they treated as independent and compared the mean difference ($A-B-B+A$) with its variance estimated from the forty-eight values in the forty-eight sandwiches, obtaining a highly significant result ($t=2{\cdot}50$).

Moreover, having demonstrated the bias in the mean of the forty-eight systematic sandwiches, they proceeded to consider the accuracy of arrangements in which each of the forty-eight sandwiches was randomized independently ($ABBA$ or $BAAB$), and also of arrangements in which each of the ninety-six pairs of half-drill strips was randomized independently (AB or BA). They concluded, *inter alia*, in the summary of their paper:

"2. Using an extensive uniformity test it is found that the arrangements randomizing either pairs or sandwiches of half-drill strips give smaller errors than the systematic arrangement advocated as more precise.

"3. As a consequence experimenters using the systematic arrangement* systematically underestimate their errors."

Gosset (1937) severely criticized this procedure, pointing out that one of the reasons for the significance attained in the systematic arrangement was the high correlation between parts of the same strip, which Barbacki & Fisher treated as independent, that the random arrangements gave more precise results because they were made up of smaller plots than the systematic arrangement, and that in any case the generalization from the results of a single trial contained in the third paragraph of the summary was unjustified.

There is some substance in these criticisms. As regards the second point, it is clear that some balanced arrangements are likely to be less accurate than others, and it may be legitimately contended that Barbacki & Fisher were comparing their random arrangements with a balanced arrangement which was not the best that could be devised, given these unit plots. Gosset does in fact suggest a balanced arrangement which compares favourably in accuracy (in this one trial) with the random arrangements. (This arrangement is discussed in § 7.)

The generalization in the third paragraph of the summary was clearly somewhat sweeping if it was intended to apply to a half-drill strip arrangement on any field. It seems reasonable to suppose, however, that all that Barbacki & Fisher had in mind was experiments on this particular field. (They had given an example of a set of six such experiments in their paper.) Actually, as will be shown later, the generalization does appear to be true of half-drill strips in general.

* Not "arrangements" as quoted by Gosset. The paragraph cannot therefore refer to any arrangement other than the half-drill strip.

In a way Gosset's first criticism is also a fair one, but it is a two-edged weapon, for it serves to emphasize how entirely arbitrary are the conventions that are ordinarily adopted in the calculation of the error of half-drill strip arrangements.

Looked at objectively the half-drill strip arrangement is really equivalent to alternate *whole* drill strips of the two varieties, with the additional defect that each variety is drilled by one-half of the drill only, so that any fault in the drill, e.g. a stopped coulter, will favour one variety at the expense of the other. The arbitrary division of each drill strip of each variety into two halves necessitates a special adaptation of the drill, and the harvesting of nearly twice the number of plots that would have to be harvested if alternate drill strips were sown. Yet though the experimenter habitually puts himself to considerable trouble to divide each drill strip lengthwise, he may not divide it transversely, for as Gosset says: "since such 'sheaf weights' may be positively correlated such a method of calculating the error is fallacious." As will be shown later in the paper, this longitudinal division of the drill strips is a hitherto unsuspected source of disturbance which tends to invalidate Gosset's method of estimating the error.

Having noted the high correlation between different parts of the same strip, Gosset examined Wiebe's results in more detail. He noticed that every eighth row, beginning with the third, gave an exceptionally high yield. He pointed out that Wiebe was using an eight-row drill (a point Barbacki & Fisher seem to have overlooked) and attributed this irregularity to some defect of the drill. He appears to have thought this provided a complete answer to Barbacki & Fisher's anomalous results. Actually, of course, the trial provides an excellent example of just that type of drilling defect which may completely vitiate the results of a half-drill strip experiment.

Gosset also failed to notice that it provides an example of the type of fertility wave or other periodic variation* which may equally vitiate the results, though it should perhaps be stressed that he was still contemplating revision of his paper at the time of his death, and it is very probable that he might have modified it considerably had he lived to see it through the press. Had he carried out a fuller analysis, he would have found that Wiebe's trial gives results which are far more unfavourable to the half-drill strip method than Barbacki & Fisher supposed. This examination is carried out in the next section.

5. Re-examination of Wiebe's uniformity trial

In view of the fact that Wiebe was using an eight-row drill, it would seem most reasonable to test the half-drill strip method with strips of four rows. This procedure has the additional advantage that it provides more numerical

* Apparently in this case due to irregularities of drilling—see additional Note at end of paper.

material. Since nearly sixteen drill widths are available there is no need to divide the rows into sections, and Gosset's main objection to Barbacki & Fisher's analysis is overcome.

TABLE I

Yield of each row of Wiebe's trial (units of 100 *g.)*

Rows							
1–16	17–32	33–48	49–64	65–80	81–96	97–112	113–125
71	63	63	69	71	75	78	74
76	66	65	70	73	80	77	72
83	72	71	76	77	84	85	81
71	62	62	67	67	73	72	70
301	263	261	282	288	312	312	297
73	59	61	65	69	74	69	67
73	67	67	75	79	80	77	73
68	61	63	69	72	74	73	67
59	56	59	65	68	70	69	67
273	243	250	274	288	298	288	274
63	59	58	66	70	70	69	65
72	65	66	72	76	75	74	71
79	72	72	79	83	80	82	78
65	62	63	67	74	74	75	67
279	258	259	284	303	299	300	281
64	61	60	65	71	71	75	70
72	69	69	76	81	81	84	[76]
67	61	66	70	76	76	72	[70]
61	61	69	73	75	78	72	[70]
264	252	264	284	303	306	303	[286]

In one respect the trial differs from a proper half-drill strip experiment: the drilling was all in one direction,* so that the inequalities noted by Gosset in the two halves of the drill will be eliminated from the results.

Table I shows the total yield of each row, and of each set of four rows, in units of 100 g. The high yields of the third and to a lesser extent the sixth row of each drill width are immediately apparent. Fictitious values have been

* This was ascertained by Neyman & Pearson (1937, p. 382).

inserted to complete the last drill width. Each of these is the mean of the other seven values in the same line of the table.*

Differences of consecutive pairs of half-drill strips (taken in the same order throughout) are shown in Table II. There is a tendency for these differences to be positive, which may be explained by differences between the two halves of the drill, referred to by Gosset. The even differences are also consistently less than the odd ones. This indicates the existence of some form of periodic variation with a period equal to two drill widths—see Note on p. 465 below. The whole situation is illustrated in Fig. 1, which shows a graph of the yields of each set of four rows.

TABLE II

Differences of half-drill strips

Rows	Diffs. of half strips	$A-B-B+A$	Rows	Diffs. of half strips	$A-B-B+A$
1–8	+28	+13	65–72	0	0
9–16	+15		73–80	0	
17–24	+20	+14	81–88	+14	+21
25–32	+ 6		89–96	− 7	
33–40	+11	+16	97–104	+24	+27
41–48	− 5		105–112	− 3	
49–56	+ 8	+ 8	113–120	+23	+28
57–64	0		121–128	− 5	
			Total	+129	+127

Whatever the causes of these irregularities, their effect on the results of the half-drill strip lay-out is disastrous. The third and sixth columns of Table II show the differences $A-B-B+A$ for each sandwich, each of these values being the difference of two consecutive values in the second or fifth column. Not one of them is negative. If they are treated as independent, we obtain the following analysis of variance.

TABLE III

Analysis of the differences $A-B-B+A$

	D.F.	Sum of squares	Mean square
Varieties	1	2016·12	2016·12
Error	7	622·88	88·98
Total	8	2639·00	

* These values were adopted in ignorance of the direction of drilling.

This gives $t=4{\cdot}76$, corresponding to a probability of about 0·002. In other words, only about one random arrangement in 500 may be expected to give results as discrepant as does the systematic arrangement in this trial.

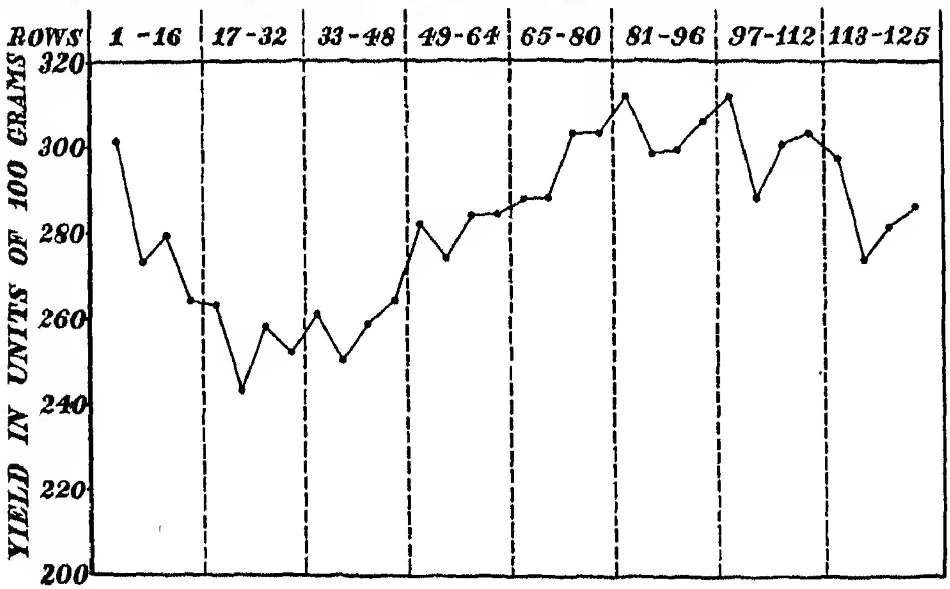
TABLE IV

Gosset's method of analysis

	D.F.	Sum of squares	Mean square
Varieties	1	1008·06	1008·06
Mean difference*	1	1040·06	1040·06
Error	14	990·88	70·78
Total	16	3039·00	

* Gosset's fertility gradient.

FIG. 1. YIELDS OF FOUR ROW MEANS IN WIEBE'S UNIFORMITY TRIAL.



Nor does Gosset's proposed method of analysis help matters. This is shown in Table IV, and is derived from the values in the second and fifth columns of Table II. The values of Table III must be divided by 2 to make them comparable with those of Table IV.

This gives a value of $t=3{\cdot}77$, corresponding to a probability of about 0·007.

6. Further points concerning the half-drill strip method

The failure of the half-drill strip method in the above example, though spectacular, might be brushed aside as exceptional. Neyman (1937), however, has quoted results obtained by Mr Sekar, which tend to show that the method is slightly less accurate on the average than is indicated by the standard error estimated by the method of Table IV, not, as Gosset supposed, more accurate.

Mr Sekar worked out values of t for 120 half-drill strip arrangements which he superimposed on different uniformity trials. (I have no particulars of what

TABLE V

Distribution of values of t in 120 half-drill strip arrangements

Limits of t	0–0·4	0·4–0·8	0·8–1·2	1·2–1·6	1·6–2·0	2·0–2·4	2·4–2·8	2·8–3·2
Nos. Expected	36·0	30·3	21·9	13·9	8·1	4·5	2·4	1·3
Nos. Observed	32	26	20	17	10	7	3	2
Discrepancy	−4·0	−4·3	−1·9	+3·1	+1·9	+2·5	+0·6	+0·7

Limits of t	3·2–3·6	3·6–4·0	4·0–4·4	4·4–4·8	4·8–5·2	5·2–5·6	5·6–∞
Nos. Expected	0·7	0·4	0·2	0·10	0·07	0·03	0·06
Nos. Observed	0	1	1	0	0	1	0
Discrepancy	+1·4						

trials were used, but it is improbable that they were all on cereals, or that those that were all provided plots that coincided with the actual half-drill strips.)

The distribution of the 120 values of t is shown in Table V.

There is a tendency to obtain too many large values of t, which, though not very marked, appears to be significant.

Why is it that Gosset's confident prediction of greater accuracy is not fulfilled, even when defects of drilling do not disturb matters? I think it is because the half-drill strip arrangement is, as already mentioned, really an arrangement in *whole* drill strips, and need not necessarily be expected to attain the full accuracy that could be obtained by using the half-drill strips in the most efficient manner, consistent with the requirements of randomization. It is very likely that randomized sandwiches, for example, are fully as accurate as systematic sandwiches, possibly even more accurate.

The question is not of great practical importance, since plant breeders rarely want to test only two varieties, and immediately the number of varieties is increased comparison by pairs, under the conditions of agricultural experimentation, becomes decidedly less efficient than the use of more comprehensive arrangements. This point is discussed elsewhere (Yates, 1935).

7. The chessboard arrangement

In place of Barbacki & Fisher's random arrangements in sandwiches and pairs Gosset proposed a new balanced arrangement of the type shown in Fig. 2,

A	*B*	*B*	*A*	*A*	*B*	*B*	*A*	.	.	.
B	*A*	*A*	*B*	*B*	*A*	*A*	*B*	.	.	.
B	*A*	*A*	*B*	*B*	*A*	*A*	*B*	.	.	.
A	*B*	*B*	*A*	*A*	*B*	*B*	*A*	.	.	.
A	*B*	*B*	*A*	*A*	*B*	*B*	*A*	.	.	.
.	.	.	.	.	.	.	.	.	.	.

Fig. 2. Gosset's balanced arrangement.

which is clearly, except at the edges, equivalent to a chessboard pattern with alternate squares (or rectangles) under the different varieties. Gosset claimed that this arrangement was likely to be more accurate, on the average, than Barbacki & Fisher's random arrangements, and in this trial the actual mean difference between the two treatments happened to be small. The result, however, appears to be largely fortuitous.

The design is of little practical importance, but it is interesting in that it provides a further illustration of the effect of arbitrarily splitting plots for the purpose of estimating the error.

It is apparent that Gosset's design can be regarded as made up of forty-eight 2×2 Latin squares

$$\begin{matrix} A & B \\ B & A \end{matrix} \quad \text{or} \quad \begin{matrix} B & A \\ A & B \end{matrix}$$

with the restriction that neighbouring squares are always of opposite type, so that his plots are really four times the area of a unit plot. Has this restriction in fact increased the accuracy over what would be obtained if the type of each Latin square were assigned at random?

The question cannot be answered with certainty from the material of a single uniformity trial, but certain indications can be obtained. Thus of the arrangements shown in Fig. 3 (all made up of 2×2 Latin squares), (1) is the

A	*B*	*B*	*A*	*A*	*B*	*B*	*A*	*A*	*B*	*A*	*B*	*A*	*B*	*A*	*B*
B	*A*	*A*	*B*	*B*	*A*	*A*	*B*	*B*	*A*	*B*	*A*	*B*	*A*	*B*	*A*
B	*A*	*A*	*B*	*A*	*B*	*B*	*A*	*B*	*A*	*B*	*A*	*A*	*B*	*A*	*B*
A	*B*	*B*	*A*	*B*	*A*	*A*	*B*	*A*	*B*	*A*	*B*	*B*	*A*	*B*	*A*
	(1)				(2)				(3)				(4)		

Fig. 3. Arrangements of 2×2 squares with varying degrees of balance.

most balanced and (4) the least balanced, (2) and (3) being intermediate. Twelve such unit arrangements of any one type can be superimposed on the

192 plots constructed from Wiebe's trial by Barbacki & Fisher. If a random choice is made for each of the twelve units between the arrangement and its complement (i.e. with A and B interchanged), an arrangement giving a valid estimate of error will result. If Gosset's arguments are correct, the use of the unit with most balance should give the lowest error. Actually the opposite is the case, as is shown in Table VI.*

TABLE VI

Variation associated with arrangements having varying degrees of balance

	D.F.	Mean square (units of a single plot)
Arrangement (1)	12	29,120
Arrangement (2)	12	10,399
Arrangement (3)	12	9,207
Arrangement (4)	12	10,915
Mean (2×2 Latin squares)	48	14,910
Randomized sandwiches	48	28,994
Randomized pairs	96	49,106

Arrangements (2), (3) and (4) all show less variation than arrangement (1), the difference between (1) and each one of the others being significant. This implies, *inter alia*, that random 2×2 Latin squares are likely to be more accurate than Gosset's more elaborately balanced arrangement.

The power of the Latin square in increasing the accuracy is here demonstrated. Neither the randomized pairs nor the randomized sandwiches considered by Barbacki & Fisher are anywhere near so effective.

8. Hudson's results

As already mentioned, the whole of Gosset's case in favour of systematic arrangements did not rest on the half-drill strip arrangement. He claimed that balanced arrangements of all types gave substantial gains in accuracy, and he put forward the results of Hudson's examination of certain uniformity trials as an example of this.

Hudson examined three uniformity trials, and his results, if taken at their face value, show a very considerable gain in accuracy with systematic arrangements. In Table VII (which also gives the main particulars of the arrangements tested) the treatment mean squares of the systematic arrangements are expressed

* There are a few minor errors in the yields given by Barbacki & Fisher (their Table I), and in their sums of squares, so that the values in the last two lines of Table VI are not in exact agreement with the values given in their analyses of variance.

as a percentage of the treatment + error mean squares of the corresponding random arrangements. In only two cases are these percentages greater than 100, and their mean is 51·7, indicating that the random arrangements are on the average giving about half the information given by the systematic arrangements.

TABLE VII

Hudson's trials

Trial	Blocks	Treatments	Plots		S.E. % per plot	Percentage information
			Rows	Length (ft.)		
Mangolds (Mercer & Hall):	20	4	3	60½	4·5	20·0
60 rows of 302½ ft.	10	4	6	60½	3·3	16·1
Unit plots:	10	4	3	121	3·9	46·0
3 rows of 30¼ ft.	8	5	3	151¼	3·8	54·8
	4	5	6	151¼	3·1	52·3
Sugar beet (Immer):	20	6	1	165	5·7	108·0
60 rows of 330 ft.	10	6	2	165	5·3	129·0
Unit plots:	10	6	1	330	4·9	58·8
1 row of 33 ft.	4	6	5	165	5·1	9·1
Potatoes (Kalamkar):	32	6	1	66	5·9	83·7
96 rows of 132 ft.	16	6	1	132	4·4	61·1
Unit plots:	16	6	2	66	6·2	41·0
1 row of 22 ft.	8	6	2	132	5·4	25·7
	8	6	4	66	5·5	65·5
	4	6	8	66	11·8	4·1

If this could be accepted as a true estimate of the average gain with systematic arrangements, it is clear that their advocates might make a strong case for their employment. The results are not very convincing, however. Only three uniformity trials are used, the plots in most of the arrangements are only one or two rows wide, and the random arrangements are in all cases randomized blocks. It is well known that Latin squares are in general substantially more accurate than randomized blocks, and examination of these systematic arrangements makes it clear that they are eliminating fertility differences in much the same way as do Latin squares. In any experiment in which the number of replicates is as great as the number of treatments one or more Latin squares would be the natural arrangement to adopt, and Hudson's comparison of his balanced arrangements with arrangements in randomized blocks is consequently of little interest. Indeed the first arrangement for the mangolds is made up of repetitions of the special type of Latin square shown in Fig. 4, and there is no conceivable reason why an experimenter using a random arrangement for four varieties on these plots should not employ a Latin square.

Balanced arrangements of the type considered by Hudson may, however, be more effective in reducing the variance between the treatment means when the number of replicates is smaller than the number of treatments, since Latin squares cannot then be used. A possible modification of the ordinary type of design in randomized blocks, based on the split-plot Latin square, which may be of use in multiple trials, and which preserves most of the "balance" of Hudson's arrangements, is considered in § 13.

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

Fig. 4. Hudson's systematic square.

9. Tedin's investigation

So far as I know, the only comprehensive investigation of the precision of any systematic arrangement which has been published was that carried out by Tedin (1931). He compared the precision of the knight's move or Knut Vik 5×5 squares, and the diagonal 5×5 squares, with that of randomized 5×5 Latin squares, using ninety-one 5×5 squares taken from eight uniformity trials. No details are given as to size and shape of plots.

The Knut Vik squares are special balanced 5×5 Latin squares in which the varieties are as evenly spaced as possible over the field. There are two such

A	*B*	*C*	*D*	*E*
D	*E*	*A*	*B*	*C*
B	*C*	*D*	*E*	*A*
E	*A*	*B*	*C*	*D*
C	*D*	*E*	*A*	*B*

Fig. 5. The Knut Vik square.

A	*B*	*C*	*D*	*E*
B	*C*	*D*	*E*	*A*
C	*D*	*E*	*A*	*B*
D	*E*	*A*	*B*	*C*
E	*A*	*B*	*C*	*D*

Fig. 6. The diagonal square.

squares, conjugate to each other, which may be applied to any set of plots. One is shown in Fig. 5. The diagonal squares are the specially simple squares, one of which is shown in Fig. 6.

Expressing the treatment mean square as a fraction of the corresponding treatment + error mean square, Tedin obtained the results shown in Table VIII.

TABLE VIII

Mean relative errors of systematic and random squares

Knut Vik squares	Square of Fig. 5	$0{\cdot}9132 \pm 0{\cdot}0599$	$0{\cdot}9120 \pm 0{\cdot}0432$
	Conjugate square	$0{\cdot}9108 \pm 0{\cdot}0622$	
Diagonal squares	Square of Fig. 6	$1{\cdot}0496 \pm 0{\cdot}0623$	$1{\cdot}0836 \pm 0{\cdot}0468$
	Conjugate square	$1{\cdot}1176 \pm 0{\cdot}0698$	
Seven random squares (the same in each trial)			$0{\cdot}9651$

The Knut Vik squares show a greater accuracy than expectation on random theory, though the gain is nothing like so striking as in Hudson's material. Nevertheless the mean fraction 0·9120 is significantly less than unity, and would indicate that the average gain in precision (though somewhat ill determined) is of the order of 10%, and is certainly less than 20%. The diagonal squares are, as might be expected, less precise than the random squares.

What does this mean to the practical agronomist or plant breeder using 5×5 Latin squares? If he uses a random arrangement in place of a Knut Vik square he will in effect be allocating two or three of the twenty-five plots to the estimation of error. Thus he may be devoting, say, 5%* of his resources to providing valid estimates of error, the elimination of unsuspected biases, and all the other advantages that accrue to random arrangements. The experience of those engaged in practical research would indicate that such an expenditure is entirely trivial in relation to the advantages gained.

Tedin's investigation applies to only one type of systematic arrangement. Obviously more comprehensive investigations could be undertaken, but it is doubtful whether they are worth while. The modern tendency in agricultural experiments is towards the greater use of factorial design, even in simple experiments involving only a few plots. Any attempt at "balancing" such designs would lead to the utmost confusion, and would greatly reduce the value of the results.

10. Where random arrangements fail

It will be apparent, on consideration of the designs discussed in the previous sections, that certain types of balanced systematic arrangements are in general likely to be more accurate than the most suitable random arrangements on the same plots, because it is impossible to introduce the same degree of local control into random arrangements while still preserving an unbiased estimate of error.

At first sight it might be thought that some improvement on ordinary Latin squares and randomized blocks should be possible. Thus the 4×4 square of Fig. 4 possesses the property that all four treatments fall in the four 2×2 squares which go to make up the larger square, and a random selection from all the 4×4 squares having this property might be made.

Unfortunately such an arrangement does not furnish a valid estimate of error, for though it is possible to eliminate the three degrees of freedom representing the contrasts of these squares, thus satisfying the least square conditions (two of the degrees of freedom are included in rows and columns, and the remaining one is orthogonal to rows and columns), the resultant estimate of error is still biased, because the condition stated in § 2 is not fulfilled. An unbiased estimate can only be obtained by making two separate estimates of

* Note that an increase of 10% in the number of plots in an experiment does not increase the work by 10%.

error, one for the contrasts (1)–(4) of Fig. 7, and the other for the contrasts (5)–(8). Clearly the second set of contrasts is likely in general to be less variable than the first.

+	+	−	−		+	+	−	−		+	−	+	−		+	−	−	+
−	−	+	+		−	−	+	+		+	−	+	−		+	−	−	+
+	+	−	−		−	−	+	+		−	+	−	+		−	+	+	−
−	−	+	+		+	+	−	−		−	+	−	+		−	+	+	−
	(1)					(2)					(3)					(4)		
+	−	+	−		+	−	+	−		+	−	−	+		+	−	−	+
−	+	−	+		−	+	−	+		−	+	+	−		−	+	+	−
+	−	+	−		−	+	−	+		+	−	−	+		−	+	+	−
−	+	−	+		+	−	+	−		−	+	+	−		+	−	−	+
	(5)					(6)					(7)					(8)		

Fig. 7. Contrasts in a 4 × 4 square with balanced corners.

Of the twelve possible patterns of treatments the four shown in Fig. 8 are such that the treatment degrees of freedom can be partitioned into two from the first group of contrasts and one from the second. For such patterns the partition of the degrees of freedom in the analysis of variance would be as in Table IX.

1	2	3	4		1	2	3	4		1	2	3	4		1	2	3	4
3	4	1	2		3	4	1	2		4	3	2	1		4	3	2	1
2	1	4	3		4	3	2	1		2	1	4	3		3	4	1	2
4	3	2	1		2	1	4	3		3	4	1	2		2	1	4	3
	(1)					(2)					(3)					(4)		

Fig. 8. Treatment patterns for a 4×4 Latin square with balanced corners.

In the remaining eight possible patterns the partition cannot be performed in this simple manner, but the expectation of the total treatment sum of squares is as before twice the error mean square (a) plus once the error mean square (b). Thus the ordinary pooled estimate of error based on 5 degrees of freedom would be biased.

TABLE IX

Analysis of variance of a 4 × 4 square with equalized corners

	Rows	3
	Columns	3
	Corners	1
First group	Treatments	2
	Error (a)	2
Second group	Treatments	1
	Error (b)	3
	Total	15

By the exclusive use of the patterns of Fig. 8, and the analysis of Table IX, an unbiased estimate of error could be obtained. The procedure has obvious

disadvantages in such a small square, but may sometimes be of value in larger squares, although an exact test of significance for the whole group of treatments will no longer be available. A similar type of arrangement, the Graeco-Latin square, has been proposed (Yates, 1937) for eliminating the bias inherent in the semi-Latin square. Split-plot and semi-Latin squares are discussed further in §13. Many of the modern devices, such as confounding in factorial design, and the quasi-factorial methods of arranging variety trials, serve a similar purpose, introducing a greater degree of local control than is provided by arrangements in ordinary randomized blocks.

Arrangements (3) and (4) of Fig. 8 are Hudson's balanced squares. These possess certain features of balance which are not possessed by squares (1) and (2): in particular every treatment occurs once at a corner. This emphasizes the inescapable fact that *some* sacrifice must be made in order to obtain a valid estimate of error, for it is only by ensuring that the component contrasts which make up the set of results shall be allotted with appropriate frequencies to both treatments and error that we can estimate the error: if one special set of contrasts believed to be more accurate than all the others is always allotted to treatments then no valid estimate of error can be possible.

11. Effect of bias in the estimate of error on tests of significance

Gosset (1936,1937) has argued at some length that the biases introduced into the tests of significance by defective estimates of error are of little consequence, or indeed are an advantage, provided the estimates of error tend to be too large. He pointed out that if the real error is decreased, and the estimate of error correspondingly increased, the ultimate outcome will be that small effects will be judged significant less frequently than they should be, but that this will be compensated for by the greater frequency with which large effects are judged significant. Pearson (1938) has given further illustrations of the same point.

In the present paper the comparison between systematic and random arrangements has been approached from the point of view of accuracy. It is perhaps worth noting that the effect on the tests of significance of biases in the estimate of error is merely equivalent to changing the level of significance.

Thus, for example, if in a series of experiments the estimates of error variance are double what they should be, the estimate x of a treatment effect will have an estimated variance s^2 which is biased by a factor 2, so that $\dfrac{x}{s/\sqrt{2}}$ will be distributed as t, i.e. x/s will be distributed as $t/\sqrt{2}$. With 11 degrees of freedom the 1% point of t is 3·106, and therefore the 1% point of $t/\sqrt{2}$ is 2·196. The 5% point of t is 2·201, and the effect on the test of significance is therefore the same as would be produced by substitution of the 1% point for the 5% point and the use of a correct estimate of error.

Consequently no new principle is introduced by Gosset's approach. The experimenter would do just as well if he admitted frankly that he believed his experiments to be decidedly more accurate than his estimates of error indicated, and allowed for this greater accuracy by the introduction of an appropriate factor (2 in the above example). He would then be at liberty to choose whatever level or levels of significance best suited his needs. Whether, of course, his choice of the numerical factor is even approximately correct remains in doubt: as we have seen, in the case of the half-drill strip arrangement, using Gosset's method of calculating the error, the factor is in reality likely to be somewhat greater than unity. The issue would, however, be clearly defined.

Actually the object of most agricultural experiments is the estimation of the magnitude of treatment effects and varietal differences, not the establishment of the existence of such effects, and the value of any estimates that are obtained is considerably increased if their standard errors are known, since fiducial limits may then be assigned to them. One has only to look through the literature of the subject to see how frequently, in the absence of such limits, theories are put forward which are in fact entirely untenable and merely serve to bring the whole of scientific agriculture into disrepute.

On the other hand, it is of course wrong to maintain (and it has in fact never been maintained) that *no* conclusions can be reached from an experiment which does not provide a valid estimate of error. Such conclusions as are reached are less objective, and are more exposed to criticism; and many of the finer points that might have been elucidated, had valid estimates of error been available, must remain matters of pure speculation.

12. MULTIPLE TRIALS

Most agricultural experiments are in fact repeated at different places and in a number of years, for it has long been realized that responses to fertilizers, differences between varieties, etc., vary substantially from year to year and place to place. In its fullest development this leads to multiple experiments, in which similar or identical trials are carried out at a considerable number of farms in the same year, and repeated in subsequent years.

Since the comparison of the different trials itself furnishes an estimate of the variation to which the results are subject, it might be considered that no estimates of error are required for such trials. The estimates of error of the individual trials, however, are still of value.

As an example of what is likely to occur in practice, we may consider the results of a set of variety trials on barley conducted in each of two years at six farms in the state of Minnesota and reported by Immer *et al.* (1934). At each farm there were three replicates of each of ten varieties arranged in randomized blocks. The interpretation of part of the results of this set of trials has been discussed in detail elsewhere (Yates & Cochran, 1938).

The combined analysis of variance published by Immer is shown in Table X.

TABLE X

Analysis of variance for twelve varietal trials

	D.F.	Mean square
Places	5	3980·31
Years	1	2541·90
Places × years	5	1261·33
Varieties	9	350·86
Varieties × places	45	80·38
Varieties × years	9	69·92 } 48·24
Varieties × places × years	45	43·90 }
Error	216	23·28

In the above table varieties × years and varieties × places × years are not significantly different, and together may be taken to provide an estimate of the variation due to changes in weather conditions and changes of field, etc., in the two years. Their combined mean square, 48·24, is quite significantly above the error mean square. The magnitude of the variance due to these causes is estimated at

$$\tfrac{1}{3}\,(48{\cdot}24-23{\cdot}28)=8{\cdot}32$$

for any one variety at one farm in a single year. This may be regarded as the effective error when we are considering the mean of a variety at any one place.

If instead of the experiment being carried out in randomized blocks some form of systematic arrangement had been used, the error being estimated as if the arrangement were in randomized blocks, some reduction in the real experimental error variance might be expected. If this was 25 %, all the mean squares of Table XII would be reduced by $\tfrac{1}{4}$ (23·28), i.e. 5·82, except the error mean square, which would be increased by $\tfrac{1}{8}$ (23·28), i.e. 2·91. The combined interactions, varieties × years and varieties × places × years would still be significant, but the magnitude of the additional variation would be estimated as

$$\tfrac{1}{3}\,(42{\cdot}42-26{\cdot}19)=5{\cdot}41,$$

i.e. it would be underestimated by about $\tfrac{1}{3}$. Set off against this is the reduction in the effective error variance of the final results. The effective error is reduced in the ratio of 48·24 to 42·42, i.e. by 12 %. Had the systematic arrangements been particularly successful and reduced the error variance by 50 % the effective error would have been reduced in the ratio of 48·24 : 36·60, i.e. by 24 %, but the apparent variance due to weather, etc., would have only one-third its true value, and would scarcely attain significance, since the estimated experimental error variance is now inflated to 29·00.

Thus by sacrificing the estimate of error we may succeed in increasing the accuracy of the final means (over a series of years) of the varietal differences at any one place, but not to the same extent as the increase in the accuracy of the individual experiments. At the same time we lose all possibility of effectively estimating the magnitude of the variation due to weather conditions, changes of field, etc. Although this does not invalidate tests of significance of the mean varietal differences (for which the effective error is estimated from varieties × years and varieties × places × years) it is a drawback which may become of importance as soon as the finer points of varietal improvement begin to be considered. Moreover the issue is further complicated by the fact that the degree of variation produced by changes of weather, etc., may differ with the different varieties, so that items in the analysis of variance such as varieties × years and varieties × places × years cannot always be regarded as homogeneous.

The lack of a proper estimate of error is also a serious disadvantage if it is necessary to increase the accuracy of the experiments, for we cannot ascertain what increase in accuracy in the varietal means at a place may be expected on, say, doubling the number of replications in each trial. In the above example, we can say immediately that this will halve the error mean square (working, now, on a two-plot basis), and will reduce all the other mean squares (on the average) by the same amount, i.e. 11·64. Thus the effective error will be reduced from 48·24 to 36·60, i.e. a reduction of 24 %. With systematic arrangements which reduced the real experimental error variance by 50 %, however, the estimated experimental error variance would, as we have seen, be inflated to 29·00, so that the estimated reduction by doubling the number of replicates would be 14·50, giving an expected reduction in effective error, from 36·60 to 22·10, of 40 %. The actual reduction, however, would only be 5·82, i.e. 16 %. Obviously in the case of the systematic arrangement the number of experiments must be increased, since there is little to be gained by increasing the number of replicates, but in the absence of a proper estimate of the additional variation due to weather, etc., the experimenter has no means of knowing this and may be seriously misled.

It may be contended that in practice the estimate of error is not likely to be so badly wrong in systematic arrangements as was suggested in the above example, and that it provides an upper limit to the true error which is sufficiently near the true error to supply all that is really required. This will occur if the gain in accuracy is itself small, but in that case systematic arrangements have little advantage over random arrangements. Nor must it be forgotten that the estimate may possibly be an underestimate, through some source of disturbance being overlooked, as in the half-drill strip arrangement.

To sum up, systematic arrangements, when used in multiple trials, do not prevent valid estimates of error and tests of significance being made for the

more important types of difference. They do, however, fail to furnish estimates of the various classes of residual variation, as distinct from experimental error, and this prevents the most effective balance being struck between number of replicates in a single trial and number of trials, apart from any interest that attaches to these classes of variation. The loss of efficiency from this cause, and the slower progress made in improving experimental designs when the errors are unknown, may well outweigh the possible immediate gain in accuracy. In general it would appear better to use some type of random design, such as the quasi-factorial or split-plot Latin square designs, which introduce additional components of balance into the arrangement, while still furnishing valid estimates of error. The split-plot Latin square design, which is of special interest for simple varietal trials such as those conducted by Immer, is discussed in the next section.

In any case it should be stressed that even if a single systematic arrangement is used it is *absolutely essential* to allocate the varieties or treatments at random to the sets of plots receiving the same treatment in this arrangement, *and to do this afresh for every trial.* Otherwise biases may be introduced and the different treatment comparisons will be subject to varying errors. These requirements are just as important if each trial consists of only a single replication. Neglect of this precaution will cast suspicion on any conclusions that may be drawn.

13. Variety trials in split-plot Latin squares

The need is sometimes felt in varietal trials carried out at a number of centres for arrangements for a moderate number of varieties involving three or

2	1	7	10	6	11	9	4	3	8	12	5
4	6	9	11	3	5	12	8	1	2	7	10
3	5	8	12	10	7	2	1	4	11	6	9

Fig. 9. Arrangement of twelve varieties in a split-plot Latin square.

four replications only. For such trials arrangements which have as a basis a Latin square with split plots may be of use.

If the varieties to be tested are divided into groups, equal in number to the proposed number of replications, the groups may be arranged in a Latin square, with randomization within the groups of plots forming the Latin square. Fig. 9 shows such an arrangement for three replications of twelve varieties. In structure it consists of a 3×3 Latin square made up of the groups of varieties (1, 2, 7, 10), (3, 5, 8, 12) and (4, 6, 9, 11), with randomization within each set of four plots.

The analysis of variance can be conducted rigorously by subdividing it into

two parts, as in the ordinary split-plot design. The partition of the degrees of freedom is shown in Table XI. If the error mean squares (*a*) and (*b*) are combined in the ratio 2 : 9 instead of 2 : 18, as would occur if the sums of squares were

TABLE XI

Partition of degrees of freedom in a split-plot Latin square

Latin square (sets of 4 varieties)	Rows	2	Within sets	Varieties	9
	Columns	2		Error (*b*)	18
	Varieties	2			
	Error (*a*)	2			

pooled, an unbiased estimate of the average error will result. Provided that the sets of four varieties are selected afresh at random for each trial the use of an average error for all comparisons is not likely to produce any serious disturbance in the analysis of a whole set of trials, though of course no exact general test of significance for a single trial is available. If such tests are required for a single trial at least four replicates in a 4×4 Latin square will be advisable. There will then be 6 degrees of freedom available for the estimation of error (*a*).

It will be seen that arrangements of this type preserve the main features of Hudson's balanced arrangements, while still permitting unbiased estimates of error to be made. As an example four such trials were superimposed on the potato trial reported by Kalamkar, and used by Hudson. Plots 22 ft. long and four rows (12 ft.) wide, with two outside rows rejected, were used. The analyses of variance of these four trials are given in Table XII.

TABLE XII

Analysis of variance of a set of four split-plot Latin squares

	D.F.	Mean squares							
		1st trial		2nd trial		3rd trial		4th trial	
Latin square:									
Rows	2	27·37		34·77		4·10		50·77	
Columns	2	537·52		991·60		10·18		0·40	
Varieties (*a*)	2	87·21	75·35	10·55	42·78	13·23	14·54	0·15	3·16
Error (*a*)	2	63·49		75·00		15·85		6·16	
Total	8	178·90		277·98		10·84		14·37	
Within Latin square:									
Varieties (*b*)	9	19·28	27·92	26·91	23·90	13·26	12·15	8·99	5·95
Error (*b*)	18	32·24		22·40		11·60		4·43	
Total	35	62·43		81·98		11·85		7·87	
Pooled:									
Varieties	11	31·63		23·94		13·25		7·38	
Error		37·92		31·96		12·37		4·74	
Incorrectly pooled error	20	35·36		27·66		12·02		4·60	

There are several points of interest in this table. The columns of the Latin squares (which correspond to 4×3 blocks of plots) account for a large part of the variance in the first two trials, but for none of the variance in the last two. In the fourth trial, but not in the others, the elimination of the rows of the Latin square (which correspond to 1×12 blocks of plots) has been effective in reducing the variance. Both the total and residual mean squares are very different in the four trials, although they are all on the same field: this is an illustration of the well-known fact that field trials, even of identical pattern and on apparently similar areas, vary greatly in their precision, an additional reason for providing an estimate of error for each trial.

The gain in precision is shown by Table XIII, which gives the varieties + error mean squares that would be obtained in randomized block experiments on the same plots, when the blocks correspond to the rows and to the columns of the Latin square respectively.

TABLE XIII

Residual mean squares for randomized blocks and split-plot Latin squares

	D.F.	1st trial	2nd trial	3rd trial	4th trial
Rows of square as blocks	33	64·56	84·84	12·32	5·27
Columns of square as blocks	33	33·63	26·85	11·95	8·33
Split-plot Latin square	—	36·54	27·33	12·58	5·44

It is clear that except in the fourth trial (which happens to be particularly accurate) the use of the columns of the square as blocks is about as effective as the split-plot design. These are in fact the most compact form of block and would probably be used by the experienced experimenter, but the result is largely fortuitous, for with plots of twice the size (4 rows $\times$ 44 ft.), and a trial occupying the same ground as the first two of the above trials, the residual mean squares are very similar in relative magnitude, having the values 271·85, 96·04 and 104·05 respectively. In this case both forms of block are equally compact, but the use of the rows as blocks gives less than half the information obtained by the use of columns or the split-plot Latin square.

It is clear that the more the Latin square component of error exceeds the other component the greater will be the inaccuracies introduced by the fact that the former is dependent on the two degrees of freedom only. In the first of these two trials the variation between the plots of the Latin square, varieties (a) + error (a), is substantially, but not excessively, above that within these plots, varieties (b) + error (b), and in the other two trials there is little difference. If these trials are representative of the type of variation ordinarily met with, it appears that the pooled estimates of error will be quite adequate for the purpose of estimating the error of the varietal means over a number of trials.

They will be somewhat less adequate for the purpose of investigating differential responses in the different trials, but even here little serious distortion of the ordinary tests is likely to result.

The results of pooling the two estimates of error by merely summing the two sums of squares are also shown in Table XII. The biases introduced will be apparent on comparison with the properly pooled estimates of error. These biases are in no case very large, but there is no advantage in using the incorrect estimate other than a slight saving in computational labour.

An obvious refinement in statistical treatment is to provide two estimates of error for each experiment, one for the comparison of varieties falling in the same group, and the other for varieties falling in different groups. The first is derived directly from error (*b*), and the second the mean of error (*a*) and error (*b*), weighted in the ratio 1 : 3. This, however, somewhat complicates the presentation of the results, and may not be worth while.

The split-plot Latin square only differs from the so-called semi-Latin square (originally suggested by Gosset under the name of "equalized randomized blocks", and independently put forward by Pitman of Tasmania, to whom the name semi-Latin square is due) in that the same groups of varieties are used for each of the Latin square plots. If this restriction is removed it is impossible to divide the analysis of variance into two parts (unless the number of replicates is sufficiently great for a Graeco-Latin square (Yates, 1937) to be used), and the resultant estimate of error is consequently biased to the extent indicated by the last line of Table XII. There is, however, the compensating advantage that the comparisons between the different varieties, and between means of groups of varieties, vary less in precision than in the case of the split-plot Latin square Thus twelve varieties can be arranged in a 3×3 square so that within the Latin square plots each variety occurs with one other variety three times, with six other varieties twice, and with four other varieties once, the variances of the differences being $\frac{2}{3}E'$, $\frac{2}{3}\left(\frac{1}{6}E+\frac{5}{6}E'\right)$ and $\frac{2}{3}\left(\frac{1}{4}E+\frac{3}{4}E'\right)$ respectively, where E and E' are the expectations of the error mean squares (*a*) and (*b*). If a split-plot Latin square is used there will be three comparisons of the first type and eight of the last for each variety. The mean variance, as before, is equal to $\frac{2}{3}\left(\frac{2}{11}E+\frac{9}{11}E'\right)$, whereas the estimate given by the analysis of variance is $\frac{2}{3}\left(\frac{1}{10}E+\frac{9}{10}E'\right)$.

In conclusion it should be emphasized that the gain in precision obtained in this example should not be taken as necessarily representative of the average gain likely to accrue under all circumstances. Much obviously depends on the shape of plot, type of crop, and other factors. A comprehensive investigation covering a representative sample of existing uniformity trials must be undertaken before it can be decided whether the gain in precision is sufficient to outweigh the statistical defects of the design. The catalogue of uniformity trials published by Cochran (1937) is likely to facilitate the selection of material suitable for such an investigation.

14. Summary

The recent claims advanced in favour of systematic arrangements by Gosset ("Student") and others are examined. The conclusion is reached that in cases where Latin square designs can be used, and in many cases where randomized blocks have to be employed, the gain in accuracy with systematic arrangements is not likely to be sufficiently great to outweigh the disadvantages to which systematic designs are subject. In particular the available evidence, though not conclusive, indicates that the half-drill strip arrangement, which Gosset particularly favoured, is likely to be somewhat *less* accurate than suitable random arrangements occupying the same plots. On the other hand, systematic arrangements may in certain cases give decidedly greater accuracy than randomized blocks, but it appears that in such cases the use of the modern devices of confounding, quasi-factorial designs, or split-plot Latin squares which are much more satisfactory statistically, are likely to give a similar gain in accuracy.

As an example the uniformity trial chosen by Barbacki & Fisher to demonstrate the defects of the half-drill strip arrangement is re-examined. It is shown that Gosset's criticisms of Barbacki & Fisher's work, though at first sight convincing, are not as conclusive as he supposed, and that in fact this particular trial provides a striking example of just those defects which have always been attributed to the half-drill strip method by its critics.

References

Barbacki, S. & Fisher, R. A. (1936). "A test of the supposed precision of systematic arrangements." *Ann. Eugen., Lond.*, **7**, 189.

Cochran, W. G. (1937). "A catalogue of uniformity trial data." *J.R. Statist. Soc. Suppl.* **4**, 233.

Gosset, W. S. (1936). "Co-operation in large scale experiments." *J.R. Statist. Soc. Suppl.* **3**, 115.

—— (1937). "Comparison between balanced and random arrangements of field plots." *Biometrika*, **29**, 363.

Immer, F. R., Hayes, H. K. & Le Roy Powers (1934). "Statistical determination of barley varietal adaptation." *J. Amer. Soc. Agron.* **26**, 403.

Neyman, J. (1937). *Lectures and Conferences on Mathematical Statistics.* Washington, D.C., Graduate School of U.S. Department of Agriculture.

Neyman, J. & Pearson, E. S. (1937). "Note on some points in 'Student's' paper on 'Comparison between balanced and random arrangements of field plots'." *Biometrika*, **29**, 380.

Pearson, E. S. (1938). "Some aspects of the problem of randomization. II. An illustration of 'Student's' inquiry into the effect of 'balancing' in agricultural experiments." *Biometrika*, **30**, 159.

Tedin, O. (1931). "The influence of systematic plot arrangement upon the estimate of error in field experiments." *J. Agric. Sci.* **21**, 191.

WIEBE, G. A. (1935). "Variation and correlation in grain among 1500 wheat nursery plots." *J. Agric. Res.* **50**, 331.
YATES, F. (1935). "Complex experiments." *J.R. Statist. Soc. Suppl.* **2**, 181.
—— (1937). *The design and analysis of factorial experiments.* Imperial Bureau of Soil Science, Harpenden.
YATES, F. & COCHRAN, W. G. (1938). "The analysis of groups of experiments." *J. Agric. Sci.* **28**, 556.

Added Note on Wiebe's Uniformity Trial

As a result of correspondence between Prof. Pearson and Dr Wiebe, which the former has kindly passed on to me, it is possible to offer an explanation of the periodic variations in yield which led to the highly significant result in the half-drill strip arrangement discussed in § 5.

It will be apparent that slight errors in directing the drill will produce unequal spacing between the last row of one drill width and the first row of the next, a point I overlooked in my discussion. If these errors are randomly distributed this will merely result in some increase of experimental error, but if they alternate in sign one variety will be favoured at the expense of the other, and a bias will result. This has occurred in the present trial.

The intended distance between each pair of drill rows was 12 in. (8 ft. between drill strips). The actual distances between the neighbouring rows of consecutive drill strips (averages of thirty-six measurements), supplied by Wiebe, are as follows:

Drill strips	Distance in.	Drill strips	Distance in.	Drill strips	Distance in.
1 and 2	10·2	6 and 7	14·2	11 and 12	11·2
2 and 3	12·4	7 and 8	11·8	12 and 13	14·0
3 and 4	11·7	8 and 9	13·8	13 and 14	11·3
4 and 5	13·4	9 and 10	12·2	14 and 15	12·9
5 and 6	10·6	10 and 11	13·1	15 and 16	12·4

Wiebe (1937) determined the effect of increase or decrease of the spacing between drills on the yield of the edge rows of each drill strip. He found that an increase of 1 in. over the normal spacing increased the yield of each of the neighbouring rows by 258 g. This, as might be expected, is slightly less than the value, 293 g., given by assuming that the yield of a row is directly proportional to the available area. Adjusting the figures of Table I, we obtain the values for the differences $A-B-B+A$ shown on the next page.

The value of t is reduced to 1·80, so that a good deal of the original excess of A over B can be attributed to drilling rather than to periodic variations in fertility. It may be noted, however, that had the centre two rows of each half-drill strip only been retained (as might reasonably be done in an actual trial in order to

eliminate both competition effects and irregularities of drilling), the value of t would still have been 2·32 (5% point = 2·36).

Original values	Adjusted values	Original values	Adjusted values
+13	+3	0	− 6
+14	+8	+21	+ 9
+16	0	+27	+16
+ 8	−3	+28	+24

The fact that the disturbance in the original results is due to a systematic error in drilling, and not to a periodic fertility wave, does not of course affect the general issue. Indeed it serves to emphasize the numerous possibilities of bias which are always present in systematic arrangements. Had the arrangement been a random one a systematic error of this kind would have produced no harmful results.

On the other hand it should be stated, in fairness to Gosset, that as a result of inspections of Dr Beaven's trials, he became aware of this particular source of bias, and drew attention to it in an addendum to his 1923 paper, where he stated that measurements on the stubble "showed not only that such inaccuracies occur, but also that they can favour one of the varieties", and added that such measurements were customarily being made to correct for this. The alternative method of rejecting certain rows entirely is probably in more common use, but it is to be noted that many descriptions of the half-drill strip method do not mention the matter, which can hardly have been regarded as a serious source of bias.

ADDITIONAL REFERENCES

"Student" (1923). "On testing varieties of cereals." *Biometrika*, **15**, 271.

Wiebe, G. A. (1937). "The error in grain yield attending misspaced wheat nursery rows and the extent of the misspacing effect." *J. Amer. Soc. Agron.* **29**, 713.

MISCELLANEA

(i) A Correction to "A Generalization of Fisher's z test"

By D. N. LAWLEY

I wish to correct and apologize for an error in my paper in the current volume of *Biometrika* (**30**, 180–7). The derivation of the distribution of v^2 in § 2 is unfortunately wrong, and thus the quantity

$$Z = \tfrac{1}{2}\log\left\{\frac{(n_2-p+1)}{n_2}\times\frac{\alpha_{ij}A'_{ij}}{p\,|A'|}\right\}$$

does not, as supposed, follow Fisher's z distribution except when $n_1 = 1$ or as an approximation when n_2 is large.

If the distribution obtained for v^2 were correct, then the quantity $u = n_1v^2/n_2$ would be distributed as χ_1^2/χ_2^2, where χ_1^2 has n_1p degrees of freedom and χ_2^2 has (n_2-p+1). We should then have

$$E(u) = \frac{n_1p}{(n_2-p-1)},$$

and
$$E(u^2) = \frac{n_1p(n_1p+2)}{(n_2-p-1)(n_2-p-3)}.$$

Thus
$$\begin{aligned}\sigma_1^2 &= E(u^2)-\{E(u)\}^2\\ &= \frac{n_1p}{(n_2-p-1)}\left\{\frac{(n_1p+2)}{(n_2-p-3)}-\frac{n_1p}{(n_2-p-1)}\right\}.\end{aligned}$$

In actual fact the distribution of u is somewhat more complicated in form, and I have been unable to obtain an explicit expression for it. It does, however, approximate to the distribution of χ_1^2/χ_2^2 when n_2 is large, and we can obtain some idea of the nature of the approximation by finding the true mean and variance and comparing them with the values given above.

Using the notation adopted before we have for the moment-generating function of u

$$\begin{aligned}M(t) &= E(e^{ut})\\ &= E\left\{\pi^{-\frac{1}{2}n_1p}\int\cdots\int \|\alpha_{ij}\|^{\frac{1}{2}n_1}\exp\left[-\tfrac{1}{2}n_1\left(\alpha_{ij}a_{ij}-\frac{A'_{ij}a_{ij}}{n_2\,|A'|}\cdot 2t\right)\right]da\right\},\end{aligned}$$

where $da = \prod_{ij} da_{ij}$ and the multiple integral is taken over the whole space.

Therefore
$$M(t) = E\left\{\frac{\|\alpha_{ij}\|}{\left\|\alpha_{ij}-\dfrac{A'_{ij}}{|A'|}\cdot\dfrac{2t}{n_2}\right\|}\right\}^{\frac{1}{2}n_1}.$$

We may suppose without loss of generality that the variances and covariances of the distribution are all unity and zero respectively, i.e. that $c_{ij} = \delta_{ij}$ (where $\delta_{ij} = 0$ when $i \neq j$, and $= 1$ when $i = j$). Then since

$$\alpha_{ij} = \frac{C_{ij}}{|C|} = \delta_{ij}$$

we shall have
$$\begin{aligned}\frac{\|\alpha_{ij}\|}{\left\|\alpha_{ij}-\dfrac{A'_{ij}}{|A'|}\cdot\dfrac{2t}{n_2}\right\|} &= \frac{\|\delta_{ij}a'_{jk}\|}{\left\|\delta_{ij}a'_{jk}-\dfrac{A'_{ij}a'_{jk}}{|A'|}\cdot\dfrac{2t}{n_2}\right\|}\\ &= \frac{\|a'_{ik}\|}{\left\|a'_{ik}-\delta_{ik}\left(\dfrac{2t}{n_2}\right)\right\|}.\end{aligned}$$

Therefore
$$M(t) = E\left\{\frac{\| a'_{ij} \|}{\left\| a'_{ij} - \delta_{ij}\left(\frac{2t}{n_2}\right)\right\|}\right\}^{\frac{1}{2}n_1}$$
$$= E\{1 - r(2t) + s(2t)^2 - \ldots\}^{-\frac{1}{2}n_1},$$

where $| A' | \, r =$ sum of all principal minors of $| A' |$ of order $(p-1)$, and $| A' | \, s =$ sum of all principal minors of $| A' |$ of order $(p-2)$. Thus
$$M(t) = 1 + \mu_1' \frac{t}{1!} + \mu_2' \frac{t^2}{2!} + \ldots,$$
where
$$\mu_1' = \tfrac{1}{2}n_1 \times 2E(r) = n_1 p E\{A'_{11}/| A' |\}$$
$$= \frac{n_1 p}{(n_2 - p - 1)},$$
and
$$\mu_2' = 8\left\{\tfrac{1}{2}n_1 E(s) + \frac{n_1(n_1+2)}{8} E(r^2)\right\}$$
$$= n_1(n_1+2)\,P + 4n_1 Q, \qquad \ldots\ldots(1)$$
putting
$$P = E(r^2),\ Q = E(s).$$

Now when $n_1 = 1$ it is known that
$$\mu_2' = E(u^2) = \frac{p(p+2)}{(n_2-p-1)\,(n_2-p-3)}.$$
Hence from (1)
$$3P + 4Q = \frac{p(p+2)}{(n_2-p-1)\,(n_2-p-3)}. \qquad \ldots\ldots(2)$$
But it may easily be proved that
$$Q = \tfrac{1}{2}p(p-1) \times \frac{1}{(n_2-p)\,(n_2-p-1)}. \qquad \ldots\ldots(3)$$
Therefore from (1), (2) and (3) we have
$$\mu_2' = E(u^2) = \frac{n_1 p(n_1 p+2)}{(n_2-p-1)\,(n_2-p-3)} - \frac{2n_1 p(n_1-1)\,(p-1)}{(n_2-p)\,(n_2-p-1)\,(n_2-p-3)}.$$
Hence
$$\sigma^2 = \mu_2' - \mu_1'^2 = \sigma_1^2 - \frac{2n_1 p(n_1-1)\,(p-1)}{(n_2-p)\,(n_2-p-1)\,(n_2-p-3)}.$$
It will be seen that although $\mu_1' = E(u) = E(\chi_1^2/\chi_2^2)$ the variance σ^2 of u is less than that of χ_1^2/χ_2^2 by an amount
$$\frac{2n_1 p(n_1-1)\,(p-1)}{(n_2-p)\,(n_2-p-1)\,(n_2-p-3)},$$
and that the proportionate error in the variance is
$$\frac{\sigma_1^2 - \sigma^2}{\sigma^2} = \frac{(n_1-1)\,(p-1)}{n_2} + O\left(\frac{1}{n_2^2}\right).$$
This would seem to indicate that if $(n_1-1)\,(p-1)/n_2$ is fairly small the error made by supposing u to be distributed as χ_1^2/χ_2^2 will not be very serious, and hence the $x\%$ point of Fisher's z with degrees of freedom $N_1 = n_1 p$ and $N_2 = (n_2 - p + 1)$ will be an approximation to the x % point of Z (defined as before on p.   above).

A better approximation may be obtained by supposing u to be distributed as the ratio of two χ^2, but altering the degrees of freedom so that the mean and variance of u have the correct values. Then χ_1^2 and χ_2^2 have degrees of freedom N_1' and N_2' respectively, where
$$N_1' = \text{nearest integer to } \{1 + (n_1-1)\,(p-1)/n_2\}\, n_1 p$$
and
$$N_2' = \text{nearest integer to } \{1 + (n_1-1)\,(p-1)/n_2\}\, (n_2 - p + 1).$$

If we now find the x % point of a z having degrees of freedom N_1' and N_2' it will be a further approximation to the x % point of Z.

To illustrate this consider the numerical example which I gave.

We had $$p = 2, \quad n_1 = 5, \quad n_2 = 30$$

and $$(n_1 - 1)(p-1)/n_2 = 2/15.$$

Thus $$N_1' = 11, \quad N_2' = 33.$$

The 0·1 % point of z with degrees of freedom 11 and 33 is 0·687 (approximately).

This is a better approximation to the 0·1 % point of Z than the value 0·728 previously obtained by taking degrees of freedom 10 and 29.

It will be noted that the significance of the value of Z obtained from the sample, i.e. $Z = 1{\cdot}0026$, is still further increased.

(ii) Twenty-Five Years of Health Progress. By L. I. Dublin and A. J. Lotka. New York: Metropolitan Life Insurance Company.

The Metropolitan Life Insurance Company of New York, the largest of its kind in the world, has just completed a mortality investigation of great social importance. During the quarter century from 1911 to 1935 the weekly premium-paying policyholders of the company, growing from eight to seventeen millions, contributed over three hundred and forty-six million years of life exposed to risk of death between the ages of 1 and 74; of these lives assured, 3,200,000 died. Such gigantic figures, collected and analysed with precision cannot fail to contain facts of great importance to all concerned in that important task, the lengthening of human life.

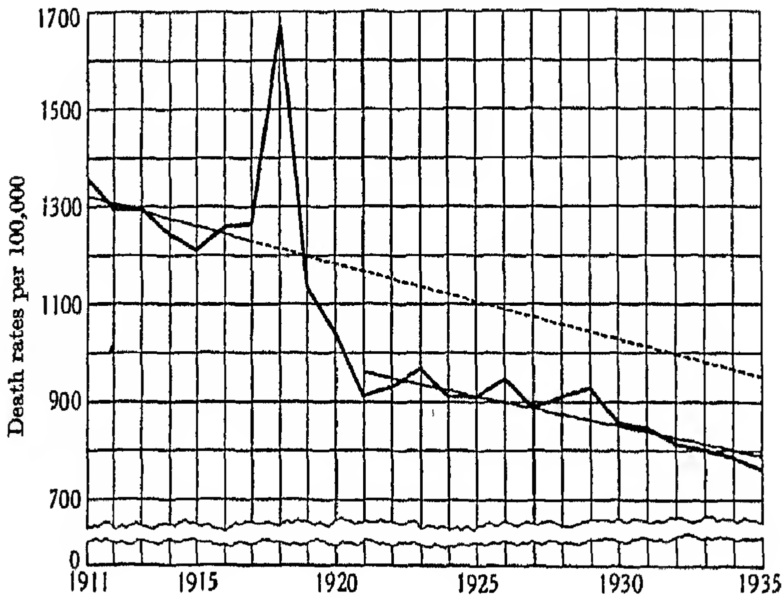
The figures relate to a fairly representative sample of the urban dwellers of the United States, and their interpretation is not immediately applicable to any assigned section of the population of England or, in particular, to the policyholders of large industrial assurance companies operating in this country. Nevertheless the broad classification of the changes in the relative importance of the various causes of death in the twenty-five year period under review, and the suggested explanations of such changes, are probably similar to the results which would be obtained in an English experience. Apart from the question of their applicability to conditions over here, the results of the initiative of the Metropolitan Life are of extraordinary interest and value, partly because they are based on such large numbers exposed to risk, partly because they are interpreted so skilfully, but mainly because they demonstrate plainly the effect of a modern environment on a body of lives resident in a civilized and progressive state.

The importance of this effect may be appreciated when it is stated that the average lifetime has been extended, during the period considered, by nearly fourteen years; or, stated another way, the standardized death rate (based on the 1901 "standard million" of England and Wales) has fallen from 1355 in 1911 to 763 per 100,000 in 1935—a decrease of nearly 44 % on the 1911 figure. The mode of this decrease is of such interest that, by kind permission of the Metropolitan Life Insurance Company, a graphical representation of the course of the standardized annual death rate per 100,000 between 1911 and 1935 is reproduced below. The curve may be analysed into three sections, 1911 to 1917, 1917 to 1921 and 1921 to 1935, the first and last with fairly uniform and almost parallel downward slopes, the second, a period of sudden and arbitrary changes. The trend line of the first period has been continued (dotted) up to 1935 and shows that immediately after the influenza pandemic there occurred an improvement in mortality which has advanced the curve of death rates by about thirteen years. The apparent explanation that the post-pandemic decline was due to the extermination, by influenza or its concomitants, of a large number of chronically invalid persons will not bear inspection, for the heavy death rate in 1918 was due to an increase in the number of deaths of young and middle-aged persons, the former being free from the degenerative diseases. The authors incline to the view that the sudden change in the level of mortality after 1918 may be due to the alteration in the bacteriological environment caused by the influenza epidemic.

After the opening chapters on the trend of longevity and the general mortality from all causes, the principal individual causes of death are dealt with, more or less in the order in which they appear in the International List of Causes of Death. In each case age, sex

All Causes of Death
Standardized Annual Death Rates per 100,000
Total Persons. Ages 1 to 74 Years

Metropolitan Life Insurance Company, Industrial Dept., 1911 to 1935



and colour (negro or white) are differentiated and the trend of the mortality rates is considered. The data and their interpretation are so rich and manifold in their implications that it would be supererogatory to comment upon them in detail. The extraordinary control of diphtheria accomplished in New York city, the striking decline in mortality from tuberculosis, the exaggerated, but commonly held, views upon the increase in cancer mortality, the desirability of further improvement in the mortality from the cardiovascular-renal diseases at the middle ages of life, the "wholesale slaughter" caused by automobile accidents, all these and many other things of more than passing interest find their place in this epic account of the amelioration in mortality rates produced by the advances in medical science and the spread of the public health movement. That the Metropolitan Life has played an important part in educating the American public to its duty in matters of health is indicated at many points of this book; that it intends to continue the education of its increasing circle of lives assured augurs well for the future of the health of the American people.

H. L. S.

(iii) Note on Professor Pitman's contribution to the theory of estimation

BY E. S. PEARSON

IT is to be hoped that Prof. Pitman's interesting contribution, published on pp. 391–421 above, will later give rise to further discussion on the problem of estimation in this *Journal* or elsewhere. There is one point on which I should, however, like to make a brief comment now. In the footnote to p. 392 Prof. Pitman suggests that his paper will show that Fisher's theory of fiducial probability and Neyman's theory of confidence intervals are essentially the same. That they are closely related, and that in very many practical cases they will lead to precisely the same form of procedure, is evident. Nevertheless, I feel that there are certain differences in the initial approach which at the present stage of development of the theory of interval estimation it is important to keep clear, since otherwise apparent disagreement arising at a later stage may lead to unnecessary misunderstanding.

I believe I am correct in saying that the following has been Prof. Neyman's line of approach to the subject. He has considered the basis of a general procedure which will provide rules for obtaining from observed data an interval that will cover the unknown parameter with a given probability. The probability is associated with repeated employment of that particular rule or method and thus if, for a specified sample of n observations, it happens that two rules lead to the same interval but associate with it different probabilities, there is no inconsistency. For example, if $x_1, x_2, \ldots, x_n$ be a random sample of n observations from a normal population, and

$$s^2 = \sum_i (x_i - \bar{x})^2/n,$$

$$w = \text{range} = \text{largest } x - \text{smallest } x,$$

it is possible to determine the multipliers a and (approximately) b so as to make the following statements about the unknown standard deviation σ in the sampled population:

$$a_1 s \leqslant \sigma \leqslant a_2 s, \text{ probability of being correct, } 0{\cdot}99; \qquad \ldots\ldots \text{(A)}$$

$$b_1 w \leqslant \sigma \leqslant b_2 w, \text{ probability of being correct, } 0{\cdot}98. \qquad \ldots\ldots \text{(B)}$$

It is then possible, if unlikely,* that the configuration of the sample x's will be such that $a_1 s = b_1 w$, $a_2 s = b_2 w$, so that two different probability statements are associated with the same interval. Following Neyman's approach, there is no inconsistency in this result, since one probability is associated with the employment of the s-rule, the other with the w-rule. It is only when we try to divorce the probability measure from the rule and to regard the former as something associated with a particular interval, that the need for a unique probability measure seems to be felt. It is such a measure, no doubt, that Fisher would define as a fiducial probability.

The following quotation from the Appendix of Neyman's paper on "Aspects of the representative method" (1934, p. 624) will illustrate this idea further. He was discussing the prediction of limits for the unknown proportion, p, of black balls in a bag after X black balls had appeared in a randomly drawn sample of three balls, and wrote:

"Having noticed this, we fix a rule as follows:

"If in the sample which we shall draw, X will have the value

$$\begin{array}{llll} X = 0 & \text{then we shall state that} & & 0 \leqslant p \leqslant \pi_1'', \\ X = 1 & ,, & ,, & \pi_2' \leqslant p \leqslant \pi_2'', \\ X = 2 & ,, & ,, & \pi_3' \leqslant p \leqslant \pi_3'', \\ X = 3 & ,, & ,, & \pi_4' \leqslant p \leqslant 1. \end{array}$$

* This will mean, in the first place, that the particular a and b factors chosen are such that $a_1/a_2 = b_1/b_2$, and then that the sample is one in which $w/s = a_1/b_1$.

We are aware that the statement which we shall make, in applying this rule to the result of actual sampling, may be wrong or may be true. We calculate the probability, P, that the statement will be a true one, and try to arrange the system of values of the π's so as to have $P \geqslant 0{\cdot}95$.... Making statements following the rules set out above, we know something important about the results of these statements: the probability that we shall be wrong is then $\leqslant 0{\cdot}05$."

We may usefully compare this statement with one from Prof. Pitman's present paper. Thus he writes (p. 396):

"The statement $$a \epsilon I(x_1, \ldots, x_n) \tag{4}$$

is a variable statement which is a function of $x_1, \ldots, x_n$. When particular, actually observed values of $x_1, \ldots, x_n$ are inserted in it, we obtain a definite statement about the unknown parameter a that is either true or false, and we shall not know which it is; but we do know that the probability that the variable statement (4), when used in this way, will give a true particular statement about a is α (supposed constant)....If we decide upon α, say 0·95, and then define I accordingly, we shall have a rule for automatically making a definite statement about the unknown parameter a whenever a set of values of the chance variable X is observed. A statistician using this rule can expect to be right about 95 times out of 100."

The correspondence between these two descriptions of the meaning of the probability statements associated with a confidence or fiducial interval is clear. The essential point of this agreement is that the probability of 0·95 is not the probability that the parameter estimated lies between any fixed limits but that a *variable statement about this parameter, made according to a specified rule, will be correct.* Having started with this common interpretation of the probability statement associated with an interval, the further steps taken by Neyman and Pitman diverge. The difference is exemplified by a sentence which I have omitted from the quotation from Pitman's paper:

"As R. A. Fisher expresses it, the fiducial probability of the variable statement

$$a \epsilon I(x_1, \ldots, x_n)$$

is α."

Now Fisher (1935, 1936) has emphasized that if a sufficient estimate of the unknown parameter a exists, a *fiducial* statement can only be made in terms of this estimate, on the grounds that it alone contains the whole of the available "information". When there is no sufficient estimate he has suggested (1936, pp. 256–7) another possible line of attack. It is this suggestion, involving the use of the sampling distribution of an estimate within samples having a given "configuration", which Pitman has followed out. It involves what is essentially a different method from Neyman's of choice between possible rules for determining an interval from given data.

It will be noted that when Prof. Pitman comes to apply his theory to the case of the normal distribution (pp. 406–8), all his fiducial statements regarding the unknown population standard deviation are expressed in terms of $S = \Sigma(x_r^2)$, that is to say, in terms of the sufficient statistic. Neyman's approach involves no initial limitation of this kind; as stated above, the interval could be defined in terms of the sample range. If the confidence limits which he accepts finally for the unknown variance are also expressed in terms of S, he has arrived at this result by a different route. Further, it is a route which, when there is no sufficient statistic, it can be shown will not always lead to the same solution as Pitman's.

It is not difficult to see just where this divergence, after initial agreement, has occurred. Neyman (1937) has shown that *any* system of confidence intervals is equivalent to some system of "regions of acceptance". Consequently, when making a choice out of an unlimited set of regions of acceptance so as to satisfy a maximum criterion as described below, he is sure of obtaining the absolute maximum. On the other hand, if I understand Prof. Pitman correctly, a restriction is placed at an early stage on the form of his regions of acceptance (p. 395); these are composed of intervals I' from the lines L which are to be such that $P\{I' \mid L\} = \alpha$ *and is constant for every* L (p. 394). Samples represented by points on a given L

all have what has been termed by Fisher the same configuration. In introducing this restriction Pitman is following Fisher's approach and not Neyman's.

It may be useful if I conclude with a brief description of what I have referred to as Neyman's maximum criterion. Having established the procedure leading to the association of a probability statement with a specified rule for determining an interval, it becomes necessary from the practical point of view to make a choice between alternative rules. Here, Neyman would say, there can be no question of an absolute right or wrong. All that can be done is to suggest a principle or principles, the following of which appears to have a strong intuitional appeal; to base an appeal on the consequences that will follow from the continued application of a given rule is a procedure which has been accepted as intuitionally sound by the human mind.

In Neyman's view, in the example of the normal curve given above, to say that because $S = \Sigma(x_i - \bar{x})^2$ and $\bar{x}$ are jointly sufficient statistics with regard to the population standard deviation and mean, σ and ξ, therefore the statement (A) is to be preferred to (B), is not by itself an argument with direct enough appeal to be convincing. To say that S contains the whole of the relevant information about σ does not provide an answer until we have been able to define just what is the nature of the information that we hope to obtain. In any case the principle of sufficiency could not be enough to determine the most appropriate interval:

(1) It will not suffice if there is no sufficient statistic. This suggests that the general principles of choice should lie somewhat deeper; they may result in the choice of a sufficient statistic when it exists, but this is a secondary result, not a primary reason.

(2) Even when it has been decided to base the rule on a sufficient statistic, we are still left in doubt as to how to select, e.g. in equation (A), from the infinite set of pairs of factors a_1 and a_2, with all of which the same probability will be associated. Again some deeper basis of choice is needed.

(3) There might well be problems in which, even when a sufficient statistic exists, the use of a rule based on some other function of the observations would have a stronger appeal. E.g. speed in calculation, inadequacy of recorded data, etc. At any rate, it is desirable to keep an open mind on such points, and allow elasticity in method.

Suppose that in the simple case of the interval estimation of a single parameter θ, we write a statement in the following form:

"There is a probability of α that, in following a specified rule for calculating from the data the limits T_1 and T_2, this statement is true:

$$T_1 \leqslant \theta \leqslant T_2."$$

Under the heading of "consequences of applying the rule", would come information on such points as:

(1) The distribution of the length of interval $T_2 - T_1$, in sampling from a population with θ fixed.

(2) The probability that values of θ differing from the true value θ_0, are included in the interval.

Neyman has suggested that the selection of the appropriate rule should be based in some way on a consideration of (2), on the grounds that an objective with a simple intuitional appeal is the following:

"If θ_0 is the true value of the parameter in the sampled population, and θ_1 some other value not equal to θ_0, then it is desirable to make the chance that the interval includes θ_1 decrease as rapidly as possible as $|\theta_1 - \theta_0|$ increases."

This approach links up with that from which Neyman and I have attacked the problem of testing statistical hypotheses, but its justification does not rest on the fact that it is so related, but rather on what I have termed its intuitional appeal. In so far as it leads to the choice of an interval based on a sufficient statistic if one exists, that is valuable knowledge. Our point of view has, however, been that no property of mathematical functions can

be accepted as the primary reason for choice of method, because such properties can hardly supply the practical experimenter with really satisfying reasons for a choice between alternatives. And if the object of the mathematical statistician is to provide tools for practical use, it seems important that the connexion between the abstract and the perceptual should be expressible in terms of the simplest possible probability concepts.

REFERENCES

FISHER, R. A. (1935). *Ann. Eugen., Lond.*, **6**, 391.
—— (1936). *Proc. Amer. Acad. Arts Sci.* **71**, 245.
NEYMAN, J. (1934). *J.R. Statist. Soc.* **97**, 558.
—— (1937). *Philos. Trans.* A, **236**, 333.